Award Number: W81XWH-08-1-0021

TITLE: Automated Neuropsychological Assessment Metrics Version 4 (ANAM4): Select Psychometric Properties and Administration Procedures

PRINCIPAL INVESTIGATOR: Susan P. Proctor D.Sc.

CONTRACTING ORGANIZATION: The Henry Jackson Foundation for the Advancement of Military Medicine, Inc.
Bethesda MD 20817

REPORT DATE: December 2015

TYPE OF REPORT: Annual

PREPARED FOR: U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE	2. REPORT TYPE	3. DATES COVERED
December 2015	Annual	01 Dec 2014 – 30 Nov 2015

4. TITLE AND SUBTITLE
Automated Neuropsychological Assessment Metrics Version 4 (ANAM4): Examination of Select Psychometric Properties and Administration Procedures

5a. CONTRACT NUMBER
W81XWH-08-1-0021

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)
Susan P. Proctor, D.Sc. and Kristin J. Heaton, Ph.D.

E-Mail: susan.p.proctor.civ@mail.mil; kristin.j.heaton.civ@mail.mil

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
The Henry Jackson Foundation for the Advancement of Military Medicine, Inc.
Bethesda MD 20817

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)
U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

10. SPONSOR/MONITOR'S ACRONYM(S)

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION / AVAILABILITY STATEMENT
Approved for Public Release; Distribution Unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT

The ability to accurately and efficiently evaluate neurocognitive status of US Warfighters exposed to diverse operational and experimental conditions is of critical importance to the ongoing mission and Force 2025 objectives of the United States military. The Automated Neuropsychological Assessment Metrics (ANAM) is a computer assisted tool for evaluating neurocognitive performance with demonstrated effectiveness for application in a wide range of military operational and research testing scenarios. The primary objective of this project is to examine select psychometric and administration properties of the ANAM4. Four studies were proposed as part of this overall effort: 1) examine common use practices and determine the effect of specific administration procedures on ANAM4 performance; 2) assess the test-retest reliability and practice effects of individual ANAM4 test modules; 3) examine the validity of the ANAM4 Mood Scale, and 4) establish a representative normative dataset of ANAM4 performance outcomes specifically for use with Army National Guard service members. Data collection for Studies 1-3 is complete; data collection for Study 4 is completed in 4 states (Minnesota, Maine, Arizona, Montana), nearing completion in Kentucky and Texas, am dos commencing in New Hampshire. The Study 4 protocol procedures are currently pending approval in Pennsylvania. Data analyses and manuscript preparation for all four studies is ongoing, with all primary data analyses completed and manuscripts nearing completion for Studies 1-3.

15. SUBJECT TERMS
ANAM, cognitive, assessment, psychometrics, validity, reliability, normative

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			USAMRMC
Unclassified	Unclassified	Unclassified	Unclassified	47	19b. TELEPHONE NUMBER *(include area code)*

Table of Contents

INTRODUCTION

The ability to accurately and efficiently evaluate neurocognitive status of U.S. warfighters under diverse operational and experimental conditions is of critical importance to the ongoing mission and *Force 2025 and Beyond* objectives of the U.S. military. The Automated Neuropsychological Assessment Metrics Version 4 (ANAM4) is a computer-assisted tool for evaluating neurocognitive performance with demonstrated efficacy for application in a broad range of military operational and research testing scenarios. The primary objective of this multi-study project is to examine select psychometric and common administration properties of the ANAM4. This project includes four studies that address different psychometric and administrative elements of the ANAM4, each critical to the understanding and utilization of this computer-assisted cognitive assessment system. Study 1 examines common use practices and their impact on ANAM4 performance. Study 2 assesses the test-retest reliability and practice effects of individual ANAM4 test modules. Study 3 examines the validity of the ANAM4 Mood Scale. Study 4 aims to establish a nationally-representative normative dataset of ANAM4 performance outcomes specifically reflecting Army National Guard Service members.

BODY

This project (which includes four studies) was funded 01 December 2007. The originally approved study timeline/SOW is presented in **Table 1.**

Table 1: Statement of Work/Study Timeline (Original, 2007)

Year 1	**Months 1-2**	**Task 1**	Plan and finalize logistics for Phase I (Studies 1-3)
	Months 3-12 (Dec 2008)	**Task 2**	Subject recruitment, data collection and data management for Studies 1-3
Year 2	**Month 13-14**	**Task 3**	Perform preliminary data analyses for Study 3
	Month 15-24 (Dec 2009)	**Task 4**	Complete data collection for Study 1
		Task 5	Perform preliminary data analyses for Study 1
		Task 6	Continue recruitment, data collection and data management for Study 2 & 3
		Task 7	Complete data collection for Study 3
Year 3	**Month 25-36 (Dec 2010)**	**Task 8**	Complete data collection for Study 2
		Task 9	Plan and finalize logistics for Phase II (modified Study 4)
		Task 10	Complete data analyses for Studies 1, 2, 3
		Task 11	Preparation of journal manuscript(s) for Studies 1, 2, 3
		Task 12	Preparation of Project report for Studies 1, 2, 3
		Task 13	Set-up data management procedures for Study 4

Table 1: Statement of Work/Study Timeline (Original, 2007) (continued)

Year 4	**Month 37-48 (Dec 2011)**	**Task 14**	Initiate data collection procedures for Study 4
		Task 15	Carry out data collection procedures for Study 4
		Task 16	Initiate integrative data management structure set up for Study 4
		Task 17	Operationalize database for Study 4 analysis scheme
		Task 18	Perform preliminary data analyses for Study 4
		Task 19	Complete data collection procedures for Study 4
Year 5	**Month 49-60 (Dec 2012)**	**Task 20**	Complete data analyses for Study 4
		Task 21	Prepare Study 4 manuscript(s) for peer review
		Task 22	Preparation of Project Final Report

A request for a 12 month no-cost extension for this study was approved on 7 November 2012, extending study activities through December 2013. A modified statement of work, approved as part of the no-cost extension, is presented in **Table 2**.

Table 2: MODIFIED SOW for remaining PROJECT Tasks and STUDY TIMETABLE (Nov 2012)

Year 4	**Month 37-48 (Dec 2011)**	**Task 14**	Initiate data collection procedures for Study 4
		Task 15	Carry out data collection procedures for Study 4
		Task 16	Initiate integrative data management structure set up for Study 4
		Task 17	Operationalize database for Study 4 analysis scheme
Year 5	**Month 49-60 (ending Dec 2012)**	**Task 18**	Conduct data collection procedures for Study 4 (cont'd)
		Task 19	Complete manuscript preparations/submissions for Studies 1-3
		Task 20	Set up/operationalize data analyses plan for Study 4
Year 6	**Month 61-72 (ending Dec 2013)**	**Task 21**	Complete data collection for Study 4
		Task 22	Complete data analyses for Study 4
		Task 23	Prepare Study 4 manuscript(s) for peer review
		Task 24	Preparation of Project Final Report

A request for a second12 month no-cost extension for this study was approved on 25 September 2013, extending study activities through December 2014. The modified statement of work is presented in **Table 3**.

Table 3. MODIFIED SOW for remaining PROJECT Tasks and STUDY TIMETABLE (Nov 2013)

Year 6	**Month 61-72 (ending Dec 2013)**	**Task 21**	Conduct data collection procedures for Study 4 (cont'd)
		Task 22	Initiate data quality control checks and preliminary analyses for Study 4.
Year 7	**Month 73-84 (ending Dec 2014)**	**Task 23**	Complete data collection for Study 4
		Task 24	Complete data analyses for Study 4
		Task 25	Prepare Study 4 manuscript(s) for peer review
		Task 26	Preparation of Project Final Report

A request for an additional 12 month no-cost extension for this study was approved on 28 October 2014, extending study activities through November 2015. The modified statement of work is presented in **Table 4**.

Table 4. MODIFIED SOW for remaining PROJECT Tasks and STUDY TIMETABLE (Oct 2014)

Year 7	**Month 73-84 (ending Dec 2014)**	**Task 23**	Initiate external data request procedures for Study 4
		Task 24	Conduct data collection procedures for Study 4 (cont'd)
		Task 25	Continue data quality control checks and preliminary analyses for Study 4 • Following each data collection trip, the newly collected data are entered into database and cleaned and preliminary data checks conducted
Year 8	**Month 85-96 (ending Dec 2015)**	**Task 26**	Complete 100% data collection goal for Study 4 (with ARNG national sample from at least 8 geographically representative US states)
		Task 27	Complete data analyses for Study 4 • With 100% data collected, complete data analyses to address Study 4 research hypotheses
		Task 28	Prepare Study 4 manuscript(s) for peer review • With completion of Study 4 analyses and manuscript preparation, travel to present findings at national conference forum is planned
		Task 29	Preparation of Project Final Report

A final request for no-cost extension, extending study activities through 31 August 2016, was approved on 30 October 2015. The complete statement of work with modified tasks for Years 7-9 (shaded) is presented in **Table 5.**

Table 5. MODIFIED SOW for remaining PROJECT Tasks and STUDY TIMETABLE (Oct 2015)

Year 1	**Months 1-2**	**Task 1**	Plan and finalize logistics for Phase I (Studies 1-3)
	Months 3-12 (Dec 2008)	**Task 2**	Subject recruitment, data collection and data management for Studies 1-3
Year 2	**Month 13-14**	**Task 3**	Perform preliminary data analyses for Study 3
	Month 15-24 (Dec 2009)	**Task 4**	Complete data collection for Study 1
		Task 5	Perform preliminary data analyses for Study 1
		Task 6	Continue recruitment, data collection and data management for Study 2 & 3
		Task 7	Complete data collection for Study 3
Year 3	**Month 25-36 (Dec 2010)**	**Task 8**	Complete data collection for Study 2
		Task 9	Plan and finalize logistics for Phase II (modified Study 4)
		Task 10	Complete data analyses for Studies 1, 2, 3
		Task 11	Preparation of journal manuscript(s) for Studies 1, 2, 3
		Task 12	Preparation of Project report for Studies 1, 2, 3
		Task 13	Set-up data management procedures for Study 4
Year 4	**Month 37-48 (Dec 2011)**	**Task 14**	Initiate data collection procedures for Study 4
		Task 15	Carry out data collection procedures for Study 4
		Task 16	Initiate integrative data management structure set up for Study 4
		Task 17	Operationalize database for Study 4 analysis scheme

Year 5	**Month 49-60 (ending Dec 2012**	**Task 18**	Conduct data collection procedures for Study 4 (cont'd)
		Task 19	Complete manuscript preparations/submissions for Studies 1-3
		Task 20	Set up/operationalize data analyses plan for Study 4
Year 6	**Month 61-72 (ending Dec 2013)**	**Task 21**	Conduct data collection procedures for Study 4 (cont'd)
		Task 22	Initiate data quality control checks and preliminary analyses for Study 4
Year 7	**Month 73-84 (ending Dec 2014)**	**Task 23**	Initiate external data request procedures for Study 4
		Task 24	Conduct data collection procedures for Study 4 (cont'd)
		Task 25	Continue data quality control checks and preliminary analyses for Study 4 • Following each data collection trip, the newly collected data are entered into database and cleaned and preliminary data checks conducted
Year 8	**Month 85-96 (ending Dec 2015)**	**Task 26**	Conduct data collection procedures for Study 4 (cont'd)
		Task 27	Continue data quality control checks and preliminary analyses for Study 4 • Following each data collection trip, the newly collected data are entered into database and cleaned and preliminary data checks conducted
Year 9	**Month 97-104 (ending Aug 2016)**	**Task 28**	Complete 100% data collection goal for Study 4 (with ARNG national sample from at least 8 geographically representative US states)
		Task 29	Complete data analyses for Study 4 • With 100% data collected, complete data analyses to address Study 4 research hypotheses
		Task 30	Prepare Study 4 manuscript(s) for peer review • With completion of Study 4 analyses and manuscript preparation, travel to present findings at national conference forum is planned
		Task 31	Preparation of Project Final Report

Task 1 (Month 1-2)
Plan and finalize logistics for Phase I (Studies 1-3) – COMPLETED
All logistical aspects for USARIEM IRB approved studies (Studies 1-3) have been confirmed. Recruitment procedures, equipment, testing facilities, and other data collection elements have been finalized and are now complete

Task 2 (Month 3-12) Subject recruitment, data collection and data management for Studies 1-3 – COMPLETED
Subject recruitment, data collection and data management efforts have been completed for Studies 1-3. Recruitment of both Human Research Volunteers and civilians participants was effective and efficient.

Task 3 (Month 13-14) Perform preliminary data analyses for Study 3– COMPLETED
All preliminary data analyses for Study 3 have been completed. Initial analyses suggested that additional participants would be necessary to explore noted differences between military and civilian participants on discrete mood measures. Thus an amendment (14 July 2009) to increase enrollment from 50 to 80 participants was submitted and approved. Data analyses have been completed on this expanded sample.

Task 4 (Month 15-24) Complete data collection for Study 1– COMPLETED
Study 1 involves the examination of common use practices and specific administration procedures (individual or group administration, practice or no practice, single session or two sessions) on ANAM4 task performances. Our recruitment goal for Study 1 was 90 participants, 30 participants per condition. Enrollment data are presented in **Table 6**.

Table 6. Study 1 Enrollment

# Participants Enrolled	90
# Participants Completed	86*

**NOTE: 15 participants completed the ANAM4 without practice test modules; 15 participants completed the ANAM4 in a group setting and 15 participants completed the ANAM4 in two administration sessions. The remaining 41 participants served as controls for these discrete administration scenarios (individual administration using practice test modules and completed in a single testing session). Thus each condition had at least 30 participants, as required.*

Task 5 (Month 15-24) Perform preliminary data analyses for Study 1 – COMPLETED
Preliminary analyses (sample characterization, demographic analyses, and preliminary group analyses) on the Study 1 data set have been completed.

Task 6 (Months 15-24) Continue recruitment, data collection and data management for Study 2 & 3 – COMPLETED
Our recruitment goal for Study 2 was 90 participants, 30 participants per condition (days 1 & 7 / days 1 & 30 / 7 consecutive day retest). Recruitment goal for Study 3 was 80 participants. Recruitment goals were reached for Studies 2 and 3 and data collection has been completed for these studies.

Task 7 (Months 15-24) Complete data collection for Study 3 – COMPLETED
Data collection for Study 3 is complete. Enrollment data are presented in **Table 7**.

Table 7. Study 3 Enrollment

# Participants Enrolled	113
# Participants Completed	77

Task 8 (Months 25-36) Complete data collection for Study 2- COMPLETED
Data collection for Study 2 has been completed. Enrollment data are presented in **Table 8**.

Table 8. Study 2 Enrollment

# Participants Enrolled	99
# Participants Completed	92

Task 9 (Months 25-36) Plan and finalize logistics for Phase II (modified Study 4) – COMPLETED
The Study 4 protocol has been reviewed and approved by USARIEM IRB and Army Human Research Protections Office (HRPO) (final approval to initiate received June 2011). Endorsement of the approved Study 4 protocol was received 20 October 2011 by National Guard Bureau (NGB) and all 8 states (Arizona, Kentucky, Maine, Minnesota, Mississippi, Montana, Oklahoma, Pennsylvania) were contacted by both NGB and USARIEM study staff. Oklahoma declined participation in September 2012. We identified Texas as a suitable replacement for Oklahoma and secured NGB endorsement for the state in October 2012.

Task 10 (Months 25-36) Complete data analyses for Studies 1, 2, 3 - COMPLETED
Preliminary data analyses have been completed for Studies 1, 2, and 3. Higher-level analyses of these data, including new ANAM Composite Score and Effort Measure analyses, have also been conducted.

Task 11 (Months 25-36) Preparation of journal manuscript(s) for Studies 1, 2, 3 – COMPLETED
Manuscripts for these studies have been prepared. Data were presented at a professional meeting (Force Health Protection, 2010).

Task 12 (Months 25-36) Preparation of project report for Studies 1, 2, 3 – COMPLETED
Project summaries and completion of Studies 1-3 were included in previous continuing review reports. Manuscripts for these studies were prepared and data were reported at a professional meeting (Force Health Protection, 2010).

Task 13 (Months 25-36) Set-up data management procedures for Study 4 - COMPLETED
Study 4 data management procedures have been established. Study 4 datasets have been created and are being populated as data are obtained from field sites. Data entry and data quality and control checks have been successfully coordinated and are ongoing with data entry procedures.

Task 14 (Months 25-36) Initiate data collection procedures for Study 4 – COMPLETED
Data collection procedures were coordinated for Arizona, Montana and Maine in 2010-2011, with data collection commencing in these three states in 2011-2012.

Task 15 (Months 37-48) Carry out data collection procedures for Study 4 – COMPLETED (See Task 18, 21, 24, & 26 for further updates)
Data collection was completed in Arizona, Maine, and Montana.

<u>Task 16 (Months 37-48)</u> Initiate integrative data management structure set up for Study 4 - COMPLETED
Databases associated with Study 4 have been created and are being populated as data are obtained and subjected to data quality and control procedures.

<u>Task 17 (Months 37-48)</u> Operationalize database for Study 4 analysis scheme – COMPLETED
Data entry has commenced and databases have been refined for analytic schemes.

<u>Task 18 (Months 49-60)</u> Conduct data collection procedures for Study 4 (cont'd) – CARRIED OUT (*See Task 21, 24, & 26 for further updates*)
Data collection procedures were completed previously in three states (AZ, ME, MT) and in a fourth state (MN) during the current reporting period. Data collection is ongoing in three states (KY, NH, TX). Coordination of TAG-level approvals has been initiated with three states (Pennsylvania, Florida and Tennessee).

<u>Task 19 (Months 49-60)</u> Complete manuscript preparations/submissions for Studies 1-3 – COMPLETED (IN PROGRESS for Submission of manuscripts to accommodate
Primary data analyses for Studies 1-3 have been completed and reported at a professional meeting (Force Health Protection, 2010) during an earlier reporting period. Manuscripts were prepared but not submitted as planned in order to include additional data being generated within the laboratory.

<u>Task 20 (Months 49-60)</u> Set up/operationalize data analyses plan for Study 4 **– COMPLETED**
Primary data analytic plan for Study 4 has been established and completed. Data were populated in the Study 4 dataset as they were collected and checked for accuracy/quality.

<u>Tasks 21 (Months 61-72)</u> Conduct data collection for Study 4 (cont'd)– CARRIED OUT
Data collection continued in three states (KY, MN, TX) in 2013. Coordination of ARNG Adjutant General-level approval to initiate data collection in New Hampshire, Pennsylvania, Florida, and Tennessee was commenced. **(*See Task 26 for current update*)**

<u>Task 22 (Months 61-72)</u> Initiate data quality control checks and preliminary analyses for Study 4 - **CARRIED OUT**
Data quality control checks and preliminary analyses were carried out as planned. **(See Task 27 & 29 for current updates)**

<u>Task 23 (Months 73-84)</u> Initiate external data request procedures for Study 4 – CARRIED OUT
An external data request (with DMDC for military service history, AFQT, and additional demographic data) was initiated and completed (October 2014) for those participants from the three states in which data collection activities were completed (AZ, MT, ME). Subsequent external data request will be made as data collection efforts with each remaining state are completed.

<u>Task 24 (Months 73-84)</u> Conduct data collection procedures for Study 4 (cont'd) – CARRIED OUT

Data collection continued in Kentucky and Texas. New Hampshire was added as an approved study site in February 2014; coordination for data collection in this state commenced. Coordination of ARNG Adjutant General-level approvals continued with Pennsylvania, Florida, & Tennessee.

<u>Task 25 (Months 73-84)</u> Continue data quality control checks and preliminary analyses for Study 4: Following each data collection trip, the newly collected data are entered into database and cleaned and preliminary data checks conducted – CARRIED OUT
Data quality control checks were carried out on an ongoing basis as data collection activities were completed at each approved site. Preliminary analyses were performed on data from three states in which data collection was completed (AZ, MT, ME) and were presented (posters) at professional conferences ***(See Appendices A & B)***.

<u>Task 26 (Months 85-96)</u> Conduct data collection procedures for Study 4 (cont'd) – CARRIED OUT
Data collection is ongoing with ARNG in three states (KY, NH, TX). We are currently coordinating TAG-level approvals with two states (Pennsylvania, Tennessee). Coordination for additional data collection trips is ongoing.

Data collection continued in Kentucky with approximately 64% of the target sample (300) for this state completed. Data collection also continued in Texas with approximately 63% of the target sample completed for the state (300). Additional trips to complete data collection in Texas, Kentucky and New Hampshire have been coordinated.

Current enrollment by state is presented in **Table 9**.

Table 9: Current Study 4 enrollment

State	# Completed
Arizona	223
Maine	248
Montana	302
Minnesota	306
Kentucky	193
Texas	193
Total	**1465**

<u>Task 27 (Months 85-96)</u> Continue data quality control checks and preliminary analyses for Study 4: Following each data collection trip, the newly collected data are entered into database and cleaned and preliminary data checks conducted – CARRIED OUT
Data quality control checks were carried out as planned. Preliminary analyses have been performed on data from three states in which data collection was completed (AZ, MT, ME). These data were presented (posters) at professional conferences ***(See Appendix A and B)***.

<u>Task 28 (Months 97-104)</u> Complete 100% data collection goal for Study 4 (with ARNG national sample from at least 8 geographically representative US states) – PENDING

<u>Task 29 (Months 97-104)</u> Complete data analyses for Study 4: With 100% data collected, complete data analyses to address Study 4 research hypotheses - PENDING

<u>Task 30 (Months 97-104)</u> Prepare Study 4 manuscript(s) for peer review: With completion of Study 4 analyses and manuscript preparation, travel to present findings at national conference forum is planned – PENDING

<u>Task 31 (Months 97-104)</u> Preparation of Project Final Report - PENDING

KEY RESEARCH ACCOMPLISHMENTS

Key research accomplishments during the current study period include:

- Progress on Study 4 data collection continued but was slower than anticipated given scheduling challenges at the ARNG-level.

- Manuscripts for Studies 1-3 were revised and refined to include additional analyses related to the ANAM Composite Score and Effort Measure metrics. Manuscripts will be finalized and submitted within the next reporting period.

- USARIEM Protocol Continuing Review was reviewed and approved by the USARIEM IRB (15 July 2015); Army HRPO acknowledgment was received on 24 September 2015.

- As described above, seven states have agreed to participate in Study 4 data collection to date and have provided ARNG Adjutant General-level approval; approvals are pending in three additional states.
    - During this reporting period, data collection activities were carried out in Texas;
    - Data collection is currently 63% complete in Texas and 64% complete in Kentucky;
    - ARNG Adjutant -level approval was secured for NH; coordination of data collection activates have commenced;
    - FL ARNG declined to participate in the study;
    - Communications with ARNG headquarters staff in two states (PA, TN) continue with approvals pending.

REPORTABLE OUTCOMES

Reportable outcomes during the current study period include:

1. Reports, manuscripts, abstracts (included as Appendices)

Proctor, S.P., Heaton, K.J., Dillon, C., Rudov, S., & Vincent, A.S. Descriptive Analyses of ANAM4 TBI Performance Among a National Sample of U.S. Army National Guard Soldiers. Poster presented at the Annual Meeting of the Association of Military Surgeons of the United States. Washington, DC, Dec. 2, 2014.

Dillon, C.C., Proctor, S.P., Vincent, A.S., & Heaton, K.J. “Demographic differences on ANAM4 TBI performance among US Army National Guard Soldiers.” Poster presented at the 123rd Annual Convention of the American Psychological Association, Toronto, Ontario, Canada, August 2015.

2. Degrees and research training opportunities

In addition to Drs. Proctor and Heaton, one doctoral-level researcher, one pre-doctoral intern, six masters-level interns and 2 Bachelor-level interns have been trained to administer the Study 4 protocol for this project.

3. Collaborative funding applications related to work supported by this award

- "Eye-Tracking Rapid Attention Computation (EYE-TRAC)" (USARIEM Protocol # H09-07; Site PI: Heaton). This project was funded as a FY08 CDMRP Advanced Technology Award to Dr. Jamshid Ghajar, Brain Trauma Foundation, New York, NY (W81XWH-08-2-0646). This study examines the efficacy of a novel visual tracking system for assessing the integrity of the attention system. The ANAM4-TBI-MIL battery was used in this study to provide cognitive performance outcomes for validation of the visual tracking paradigm. Healthy military volunteers were subjected to a 26-hour period of sleep loss during which cognitive and visual tracking performance were evaluated. Test-retest reliability of the ANAM4-TBI-MIL was examined across a 2 week interval and sensitivity of the ANAM4 TBI battery to central fatigue were determined. One paper (pending) and one abstract (accepted) involve ANAM4-TBI-MIL data collected from this study:

  Heaton, K.J., Laufer, A.S., Maule, A., Vincent, A.S. Effects of acute sleep deprivation on ANAM4 TBI Battery performance in healthy US Army Service Members. *In preparation*

  Heaton, K.J., Laufer, A.S., Maule, A., Vincent, A.S.. Effects of acute sleep deprivation on ANAM4 TBI Battery performance in healthy US Army Service Members. Poster presented at the 123rd Annual Convention of the American Psychological Association, Toronto, Ontario, Canada, August 2015.

- "An Investigation of the Effects of Head Impacts Sustained during Collegiate Boxing Participation on Central and Peripheral Nervous System Function" (USAFA Protocol # FAC2007010H, PI: MAJ Brandon Doan, USAFA), was funded in part by an AMEDD Advanced Medical Technology Initiative (AAMTI) award to Dr. Heaton. In this study, the effects of mild, repetitive head impacts sustained during amateur boxing training bouts on cognitive performance outcomes were examined using the ANAM4-TBI-MIL and IMPACT cognitive test batteries. One manuscript is being re-submitted for review related to this work:

  Heaton KJ, Adam GE, Butler MA, Self B, Brininger T, Wile A, Rudolph KA, Doan B. Mild Repetitive Head Impacts and Neurocognitive Performance in Amateur Military Boxers. *Submitted to Archives of Clinical Neuropsychology.*

- "Identifying biomarkers that distinguish post-traumatic stress disorder and mild traumatic brain injury using advanced magnetic resonance spectroscopy," was funded via a Department of Defense Congressionally Directed Medical Research Programs Psychological Health/Traumatic Brain Injury (PH/TBI) Research Program award to Dr. Alex Lin, Brigham and Women's Hospital, Boston, MA. Dr. Heaton is a co-Investigator and site PI on this

project. This study proposes a multi-parametric approach using major advances on spectroscopic methods and neuroimaging to identify biomarkers that can be used to distinguish between post-traumatic stress disorder, traumatic brain injury, and their co-occurrence. This will be achieved in part by correlating quantitative MR spectroscopy results with behavioral and neuropsychological metrics (including ANAM4TBI) using newly developed algorithmic approaches that are capable of revealing discriminating metabolic markers in MR spectroscopy measurements. Data collection for this project is ongoing. Four abstracts (accepted) involve ANAM4-TBI-MIL data collected from this study:

H Liao, K Heaton, P Merugumala1, J Saurman, X Long, I Orlovsky, S Merugumala, K Rudolph, N Murphy, B Rowland, AP Lin. "Reduced NAA and Glutamate in Healthy Military Subjects Compared to Civilian Controls." Presented as a poster at the International Society for Magnetic Resonance in Medicine (ISMRM), Toronto, Canada, May 30-June 5, 2015.

Mariano L, Irvine J, Rowland B, Heaton K, Lin A. "Psychological Health/Post Traumatic Stress Disorder - Biomarkers Discovery for PTSD and mTBI using Magnetic Resonance Spectroscopy." Presented as a poster at the Military Health System Research Symposium (MHSRS). Florida, August 2015.

AP Lin, H Liao, J Saurman, P Merugumala, I Orlovsky, X Long, S Merugumala, K Rudolph, B Rowland, K Heaton "Differences in Brain Biochemistry between Military and Civilian Controls.". Presented as a poster at the Military Health System Research Symposium (MHSRS). Florida, August 2015.

Mariano L, Irvine J, Rowland B, Liao V, Ladner J, Heaton K, Lin A. "Novel Processing of Magnetic Resonance Spectroscopy Signal Enables Biomarker Discovery for PTSD and mTBI" Poster presented at AMSUS, 30 Nov - 1 Dec 2015, San Antonio, TX

- "Multimodal Assessment of Cognitive Readiness and Recovery: Initial Modeling of Physiological and Neurological Inputs" (USARIEM Protocol 15-05HC; PI: Heaton), was funded by Defense Health Program (DHPe, RDT&E, Operational Performance Sustainment; "Multimodal Assessment of Cognitive Readiness and Recovery: Modeling and Analysis of Physiological and Neurological Inputs") to Dr. Heaton and MIT Lincoln Laboratory investigator, Dr. Thomas Quatieri. This study will examine the sensitivity of a multi-modal platform for detecting change in cognitive functioning under different cognitive load conditions. The platform consists of vocal, facial, physiological (heart rate, skin conductance, respiration), and cognitive data inputs. The ANAM4 is included in the cognitive test battery. This protocol is currently under review. Three abstracts and a paper (accepted) involve ANAM4-TBI-MIL data collected from this study:

  Quatieri TF, Williamson JR, Smalt CJ, Helfer BS, Patel T, Perricone J, Ciccarelli G, Mehta DD, Ricke D, Malyska N, Palmer J, Heaton K, Eddy M, Moran J. "Comparison of vocal and EEG biomarkers to discriminate cognitive load in a working memory task." Body-Sensor-Network, MIT Media Lab, Cambridge MA, 9-11 June 2015. Poster

  Quatieri TF, Williamson JR, Smalt CJ, Patel T, Perricone J, Mehta DD, Helfer BS, Ciccarelli G, Ricke D, Malyska N, Palmer J, Heaton K, Eddy M, Moran J. "Vocal

biomarkers to discriminate cognitive load in a working memory task." Interspeech, Germany, September 2015. Presentation and paper.

Horwitz-Martin R, Quatieri T, Helfer B, Williamson J, Vian T, Lacirignola J, Shenk T, Talavage T, Palmer J, Heaton K. "Phone durations as predictors of preclinical; mild traumatic brain injury symptom severity." Poster presented at the 5th Annual Traumatic Brain Injury Conference, Washington, DC, April 15-16, 2015.

4. Related projects and collaborations initiated

- "Analyses of ANAM4™TBI Predeployment Assessment Data: USARIEM-OTSG Research Collaborative" (USARIEM #11-07HC; PI: Proctor)involves the creation of a research database system (ANAM4TBI Military Performance Database (AMP-D)) which incorporates all mandated pre-deployment ANAM4TBI assessment data from DoD military personnel (maintained by the Office of the Surgeon General, ANAM Program Office). We have initiated the process of linking these neurocognitive data with individual military service, demographic, and injury and clinical disease histories. At the conclusion of Study 4, we plan utilize the AMP-D to make comparisons between Army Active Duty and National Guard groups and examine the role of deployment-related factors on neurocognitive health and performance. A manuscript detailing the AMP-D and population demographics was submitted and has been accepted for publication:

  Proctor SP, Nieto K, Heaton KJ, Dillon CC, Schlegel RE, Russell ML, Vincent AS. 2015. Performance and Prior Injury among U.S. Department of Defense Military Personnel. *Military Medicine*, 180, 6:660-669.

- "Validation of Select Neurobehavioral Assessments for Concussion/Mild Traumatic Brain Injury (MTBI)" (USARIEM #H09-08), was intramurally funded (MRMC RAD3) to Drs. Proctor and Heaton (co-PIs). This study seeks to validate the ANAM4TBI Battery against a standard neuropsychological screening battery for mild traumatic brain injury. Data collection for this project has been completed; data analyses and manuscript preparation are underway.

- "Multidimensional MR Imaging to Assess Subtle Brain Changes Associated with Persistent Postconcussive Symptoms (PPCS) Following Mild Traumatic Brain Injury" (USARIEM Protocol #11-15-HC; PI: Palumbo, co-I: Heaton), was intramurally funded (MRMC RAD3) to Dr. Palumbo (co-I: Heaton). This study examines neuropathological changes associated with PPCS following mTBI using multidimensional magnetic resonance imaging (MRI) to determine the independent and synergistic effects of structure, function, connectivity and blood flow of the brain in subjects with mTBI. ANAM4-TBI-MIL is being used in this study to examine cognitive performance outcomes. Data collection for this study has been completed; data analyses and manuscript preparation are underway.

CONCLUSION

Analyses of data from Studies 1-3 have been completed and manuscripts are currently being revised for submission. Our results (reported in conference proceedings included in the 2010 Annual Report for this project) provide evidence supporting the Automated Neuropsychological Assessment

Metrics Version 4 (ANAM4) as a reliable and valid measure of cognitive performance under diverse administration scenarios.

Development of a nationally-representative normative dataset of Army National Guard service members' ANAM4 performance outcomes (Study 4) is currently pending completion of data collection. Preliminary results have been presented at professional conferences. The target reference dataset is intended to complement existing normative data by focusing on a subset of the general military population that research has shown differs on key demographic elements (e.g., dual career status, average age, marital/family status, and education) relative to other military components (e.g., Active Duty), and as such, is expected to facilitate the interpretation of individual National Guard service members' performance on ANAM4 tests.

Together, results from all four studies in this project will add to ongoing efforts to develop and validate the ANAM4 (and ANAM4 Military Traumatic Brain Injury Battery) as an accurate, reliable, and objective measure of military service members' cognitive performance.

APPENDIX

Appendix A: Proctor, S.P., Heaton, K.J., Dillon, C., Rudov, S., & Vincent, A.S. (2014). "Descriptive Analyses of ANAM4 TBI Performance Among a National Sample of U.S. Army National Guard Soldiers." Poster presented at the Annual Meeting of the Association of Military Surgeons of the United States. Washington, DC, Dec. 2, 2014.

Appendix B: Dillon, C.C., Proctor, S.P., Vincent, A.S., & Heaton, K.J.. "Demographic differences on ANAM4 TBI performance among US Army National Guard Soldiers." Poster presented at the 123rd Annual Convention of the American Psychological Association, Toronto, Ontario, Canada, August 2015.

Appendix C: Heaton, K.J., Laufer, A.S., Maule, A., Vincent, A.S. "Effects of acute sleep deprivation on ANAM4 TBI Battery performance in healthy US Army Service Members." Poster presented at the 123rd Annual Convention of the American Psychological Association, Toronto, Ontario, Canada, August 2015.

Appendix D: H Liao, K Heaton, P Merugumala1, J Saurman, X Long, I Orlovsky, S Merugumala, K Rudolph, N Murphy, B Rowland, AP Lin. "Reduced NAA and Glutamate in Healthy Military Subjects Compared to Civilian Controls." Presented as a poster at the International Society for Magnetic Resonance in Medicine (ISMRM), Toronto, Canada, May 30-June 5, 2015.

Appendix E: Mariano L, Irvine J, Rowland B, Heaton K, Lin A. "Psychological Health/Post Traumatic Stress Disorder - Biomarkers Discovery for PTSD and mTBI using Magnetic Resonance Spectroscopy." Presented as a poster at the Military Health System Research Symposium (MHSRS). Florida, August 2015.

Appendix F: AP Lin, H Liao, J Saurman, P Merugumala, I Orlovsky, X Long, S Merugumala, K Rudolph, B Rowland, K Heaton. "Differences in Brain Biochemistry between Military and Civilian Controls." Presented as a poster at the Military Health System Research Symposium (MHSRS). Florida, August 2015.

Appendix G: Mariano L, Irvine J, Rowland B, Liao V, Ladner J, Heaton K, Lin A. "Novel Processing of Magnetic Resonance Spectroscopy Signal Enables Biomarker Discovery for PTSD and mTBI" Poster presented at AMSUS, 30 Nov - 1 Dec 2015, San Antonio, TX.

Appendix H: Quatieri TF, Williamson JR, Smalt CJ, Helfer BS, Patel T, Perricone J, Ciccarelli G, Mehta DD, Ricke D, Malyska N, Palmer J, Heaton K, Eddy M, Moran J. "Comparison of vocal and EEG biomarkers to discriminate cognitive load in a working memory task." Poster presented at the Body-Sensor-Network, MIT Media Lab, Cambridge MA, 9-11 June 2015.

Appendix I: Quatieri TF, Williamson JR, Smalt CJ, Patel T, Perricone J, Mehta DD, Helfer BS, Ciccarelli G, Ricke D, Malysk18a N, Palmer J, Heaton K, Eddy M, Moran J. "Vocal biomarkers to discriminate cognitive load in a working memory task." Presentation at Interspeech, Germany, September 2015. Paper accepted for publication.

APEENDIX J: Horwitz-Martin R, Quatieri T, Helfer B, Williamson J, Vian T, Lacirignola J, Shenk T, Talavage T, Palmer J, Heaton K. “Phone durations as predictors of preclinical; mild traumatic brain injury symptom severity.” Poster presented at the 5th Annual Traumatic Brain Injury Conference, Washington, DC, April 15-16, 2015.

Appendix K: Proctor SP, Nieto K, Heaton KJ, Dillon CC, Schlegel RE, Russell ML, Vincent AS. 2015. Performance and Prior Injury among U.S. Department of Defense Military Personnel. *Military Medicine,* 180, 6:660-669.

APPENDIX A

Proctor, S.P., Heaton, K.J., Dillon, C., Rudov, S., & Vincent, A.S.. Descriptive Analyses of ANAM4 TBI Performance Among a National Sample of U.S. Army National Guard Soldiers. Poster presented at the Annual Meeting of the Association of Military Surgeons of the United States. Washington, DC, Dec. 2, 2014.

ABSTRACT

Limited research has focused on the neurological health and performance of U.S. Army National Guard (ARNG) personnel. In light of the dual-job occupational histories and demographic differences (i.e., older, more years of education) of ARNG compared to their Active Duty (AD) counterparts, it is important to identify and characterize possible performance differences on measures of cognitive function.

Current efforts are underway to develop a national reference sample of ARNG Soldiers' performance on the Automated Neuropsychological Assessment Metrics (version 4) TBI Military (ANAM4 TBI-MIL) battery. This reference sample will be comprised of data from a representative sample of 2,400 ARNG Soldiers from 8-10 U.S. states.

Descriptive analyses of questionnaire and performance data (n=695) from three states completed to date (Montana, Maine, and Arizona) were performed. The ARNG sample was 15% female and 30.6 (SD=9.1) years old on average; the majority (64%) had completed education beyond the high school level. ANAM4 TBI-MIL task performance was compared to published normative data from AD personnel (10% female and mean age 27.4 (SD=7.4) years). Overall, no significant performance differences were observed between the ARNG and AD on tasks involving visual memory and complex attention, while ARNG personnel performed with significantly reduced efficiency ($p<.001$) on tasks of simple attention and psychomotor speed. When comparative analyses were restricted to those 21-25 years of age, no significant differences in performance were observed.

In conclusion, neurocognitive performance differences between AD and ARNG were observed on certain neurocognitive tasks, however, results suggest these are related to demographic factors (i.e., age).

APPENDIX B

Dillon, C., Proctor, S.P., Vincent, A.S., & Heaton, K.J.
Demographic differences on ANAM4 TBI performance among US Army National Guard Soldiers. Submitted for Poster Presentation at the 123rd Annual Convention of the American Psychological Association, Toronto, Ontario, Canada, August 2015.

ABSTRACT

Several studies have examined the neurocognitive performance of the U.S. military, particularly Active Duty personnel. However, minimal research has focused on the neurocognitive performance of U.S. Army National Guard (ARNG) Soldiers. Known demographic differences between Active Duty and Reserve/National Guard personnel on such factors as age and education level may influence neurocognitive proficiencies. Thus, the goal of this analytic study was to examine the role of demographic factors on neurocognitive test performance within a multi-state cohort of ARNG personnel.

The Automated Neuropsychological Assessment Metrics (version 4) TBI Military (ANAM4 TBI-MIL) battery was developed to assess general cognitive functioning, specifically following injuries to the head. A normative dataset for the ANAM4 TBI-MIL has been created for use with U.S. Active Duty personnel. Comparable reference data are not currently available for use with Army National Guard personnel specifically. Use of appropriate reference data is critical to the accurate interpretation of test performance. Data collection from a sample of ARNG personnel designed to be representative of the current U. S. ARNG population is ongoing and upon completion will include ANAM4 TBI-MIL performance data from approximately 2,400 ARNG Soldiers from 8-10 U.S. states.

Performance data were analyzed from three states completed to date (Arizona, Maine, and Montana; n=695). The ARNG sample was 15% female and 30.6 (SD=9.1) years old on average; the majority (64%) had completed some education beyond the high school level. Significant performance differences were observed between age groups (18-24 years old; 25-34 years old; 35 years and older), with younger participants performing better on tasks measuring sustained attention, reaction time, processing efficiency, visuospatial working memory and delayed memory ($p<.001$). There was a significant benefit of advanced education (high school or equivalent vs. greater than high school) on a one test measuring basic computational skills and processing speed ($p<.001$). This benefit is not associated or confounded by age. There were no observed differences in task performance between male and female participants.

In conclusion, neurocognitive performance differences on the ANAM4 TBI-MIL battery were associated with age. However, minimal to no performance differences related to education and gender were observed. Further evaluation of demographic factors will be conducted with the complete multi-state cohort of ARNG personnel.

APPENDIX C

Heaton, K.J., Laufer, A.S., Maule, A., Vincent, A.S. (abstract submitted). Effects of acute sleep deprivation on ANAM4 TBI Battery performance in healthy US Army Service Members. Submitted for Poster Presentation at the 123rd Annual Convention of the American Psychological Association, Toronto, Ontario, Canada, August 2015.

Introduction: The Automated Neuropsychological Assessment Metrics (version 4) Traumatic Brain Injury Battery for the Military (ANAM4 TBI-MIL) is currently being used within the U.S. Army as part of a comprehensive brain injury/concussion screening program, providing a broad measure of cognitive function to aid clinicians in the assessment and treatment of brain injuries. Numerous factors endemic to military operational and training environments, including physical and mental fatigue, have been shown to produce shifts in cognitive status and mood in prior research involving military and civilian populations. Thus, the presence of these factors may confound the interpretation of cognitive performance. Although the effects of sleep loss on cognitive function have been examined in earlier versions of the ANAM, the impact of sleep loss on ANAM4 TBI-MIL battery performance outcomes has not yet been reported. Understanding the influence of factors such as fatigue on ANAM4 TBI-MIL performance is critical for accurate interpretation of test results in military service members. The impact of fatigue is also an important component of injury prevention and ensuring optimal performance and mission readiness of military service members.

Methods: The effects of acute (26 hours)sleep deprivation on cognitive performance as evaluated by the ANAM4 TBI-MIL battery were examined in 87 healthy US Army service members (68 men, 19 women), ranging in age from 18-33 with an average of 12.5 years of education. The ANAM TBI-MIL battery consists of a sleepiness scale, a mood scale and 7 additional test modules assessing reaction time, memory, processing efficiency, working memory, basic computational skills and attention. Participants completed the ANAM4 TBI-MIL battery three times during the sleep deprivation period: initial waking (baseline), ~20 hours awake, and ~26 hours awake.

Results: Across the 26 hour period of sleep loss, participants demonstrated increasingly slowed response times on 5 of the 7 cognitive test modules, including tasks of simple response speed, visual memory, working memory, processing efficiency and attention (p-values ranging from .014 to < .000). Degraded accuracy was observed on 3 of the 7 cognitive test modules, including working memory, processing efficiency, and visual memory tasks) (p-values < .000). In addition, participants reported an increase in sleepiness, a decrease in vigor and happiness and increased levels of restlessness, anxiety, anger/irritability and depressed affect (p values ranging from .002 to < .000).

Conclusions: Consistent with prior research involving ANAM and other cognitive assessment tools, results show degraded response speed and accuracy across most test modules of the ANAM4 TBI-MIL battery following a period of acute sleep deprivation. These findings provide evidence of the sensitivity of the ANAM4 TBI-MIL battery to the effects of acute sleep deprivation, an important consideration when evaluating service members in operational settings.

The views expressed in this presentation are those of the authors and do not reflect the official policy of the Department of the Army or the Department of Defense.

APPENDIX D

"Reduced NAA and Glutamate in Healthy Military Subjects Compared to Civilian Controls." H Liao, K Heaton, P Merugumala1, J Saurman, X Long, I Orlovsky, S Merugumala, K Rudolph, N Murphy, B Rowland, AP Lin. Presented as a poster at the International Society for Magnetic Resonance in Medicine (ISMRM), Toronto, Canada, May 30-June 5, 2015.

4035

Reduced NAA and Glutamate in Healthy Military Subjects Compared to Civilian Controls

Huijun Liao[1], Kristin Heaton[2], Praveen Merugumala[1], Jessica Saurman[2], Xi Long[1], Irina Orlovsky[2], Sai Merugumala[1], Kelly Rudolph[2], Nicole Murphy[2], Benjamin Rowland[1], and Alexander P. Lin[1]

[1]Center for Clinical Spectroscopy, Brigham and Women's Hospital, Boston, MA, United States, [2]Military Performance Division, US Army Research Institute ofEnvironmental Medicine, Natick, MA, United States

TARGET AUDIENCE: Researchers and clinicians with interest in brain metabolism in military medicine

PURPOSE: Many studies have examined traumatic brain injury and post-traumatic stress disorder among other neurological disorders in military subjects. A few of these research studies had used healthy civilian subjects as a control group and found significant differences between patients and controls. However, comparing civilian controls alone with military patients might introduce flaws to data analysis since there may be inherent differences between military and civilian subjects. To our knowledge, there has not been a systematic study that challenges the assumptions that the cohorts are the same. In this ^{1}H MRS study, the main objective was to investigate the validity of this assumption by detecting the significant difference in MRS quantifiable metabolites between healthy military subjects and civilian subjects.

METHODS: *Participants.* 9 healthy military subjects (including service members and veterans, mean age 32.1±9.7, 3 female, 6 male) and 9 age- and gender-matched healthy civilian controls (mean age 32.7±11.6) were recruited and consented under local IRB approval. Both healthy military and civilian subjects had no history of neurological disorders, psychological disorders or drug addiction by self-report. All subjects also underwent neuropsychological evaluation including Rivermead Post-concussive Symptoms Questionnaire, Post Traumatic Stress Disorder (PTSD) Checklist – Civilian Version, Beck Depression Inventory II, Automated Neuropsychological Assessment Metrics – version 4- TBI Battery, Wechsler Memory Scale – III Spatial Span Test, Rey Auditory Verbal Learning Test, Test of Memory Malingering, Trail Making Test - A&B, Wechsler Adult Intelligence Scale III, Processing Speed Index.

MRS data acquisition and analysis. This study was performed in a Siemens 3T MAGNETOM Skyra scanner and using 32- channel head coil. Single Voxel MRS was acquired using conventional PRESS in four different brain regions shown in Figure 1: Posterior Cingulate Gyrus (PCG; 20x20x20mm), Posterior White Matter (PWM; 20x20x20mm), Anterior Cingulate Gyrus (ACG; 20x20x20mm) and Lefttemporal Lobe (hippocampus area, Left-temp; 20x15x15mm). All voxels were acquired using TE = 30 ms, TR = 2 s, bandwidth = 1.2 kHz, 1024 complex data points, water saturation, and 128 averaged acquisitions. Unsuppressed water spectrum with the same parameters but without water suppression and 16 averages. Total scan time: 5.13 minutes per voxel. PRESS data was analyzed using LCmodel. Metabolite concentrations were expressed in institutional units as well as a ratio of metabolite to total creatine (Cr+PCr).

RESULTS: Figure 1 shows an example of a 3T spectrum acquired in a healthy military control. Among all LC-model quantifiable metabolites, glutamate and NAA concentration showed significant differences between healthy military and civilian subjects. Compared to civilian subjects, lower Glu/Cr+PCr ratios were observed in military subjects in all four voxel locations (Figure 2a) and significantly in PCG ($p<0.05$) and PWM ($p<0.001$). In addition, reduced NAA+NAAG/Cr+PCr ratios were also observed in military subjects across all four voxel locations (Figure 2b) and significantly lower in PCG ($p<0.05$) and, PWM ($p<0.001$). Cr+PCr was not found to be significantly different. All healthy civilian and military subjects were negative for post-concussive symptoms, PTSD, and depression. There were no significant differences between the two groups in their performance on neuropsychological testing.

DISCUSSION: Glutamate and NAA showed similar trends which both had lower mean ratios in the military group across all four voxel locations and the most significant reduced mean ratios in PWM. Even though similar findings were shown in glutamate and NAA, they were not highly correlated with each other with R_2=0.54 in PWM. Reduced Glu and NAA have been found in depression[1] however both cohorts did not show depression as evaluated by BDI. Regarding lower NAA, a study showed that changes of NAA may be due to different education levels[2], but we did not find significant difference in years of education between civilian and military group in this study. Therefore, the reason for lower Glutamate and NAA in healthy military subjects than civilian subjects in this study is still unclear. Future studies will include a larger cohort and additional measures to compare the two groups.

CONCLUSION: Lower Glutamate and NAA concentration in healthy military group compared to healthy civilian group indicates a difference between the two and the assumption that the two groups are the same is not true. Military studies should utilize healthy controls with similar military background.

REFERENCES: 1. Merugumala et al. MRI in Psychiatry. (2014) 2.Glodzik L et al. Psychiatry Research:Neuroimaging. 204 (1) 49–54 (2012)

ACKNOWLEDGEMENTS: This study was funded by DOD CDMRP WX81-XWH-10-1-0835. Opinions, interpretations, conclusions, and recommendations are those of the authors and are not necessarily endorsed by the United States Government.

APPENDIX E

"Psychological Health/Post Traumatic Stress Disorder - Biomarkers Discovery for PTSD and mTBI using Magnetic Resonance Spectroscopy." Mariano L, Irvine J, Rowland B, Heaton K, Lin A. Presented as a poster at the Military Health System Research Symposium (MHSRS). Florida, August 2015.

Abstract:

Background:

Post Traumatic Stress Disorder (PTSD) and mild Traumatic Brain Injury (mTBI) affect returning soldiers from Operation Iraqi Freedom and Enduring Freedom (OIF / OEF) at an alarming rate. Our study focuses on magnetic resonance spectroscopy (MRS) measurements to distinguish subjects having mTBI, PTSD, or both, with the goal of developing biomarkers from the MRS data. The assessment of metabolite concentrations in the brain is critical to understanding neurological disorders. MRS provides a non-invasive *in vivo* technique for measuring these metabolites.

Methods:

Using a simple factorial design for the experiment, military subjects fall into four categories: controls, PTSD only, mTBI only, and both PTSD and mTBI. Acquisition of the MRS data was performed on a Siemens Verio 3T scanner using a 32 channel head coil. Data was extracted in the Siemens 'twix' format which contains individual free induction decays for each channel and average. Channel weightings for the raw data were determined from the water reference signal using a singular value decomposition method designed to maximize SNR, then applied to the main data. Features were extracted from each signal using two approaches: wavelet decomposition, and LCModel. Following the designed protocol, we compared MRS features from subjects with PTSD, mTBI, or both, against controls.

Results:

MR spectroscopy signals of the brain are modeled as a superposition of the resonances from the underlying metabolites, plus distortions arising from the data acquisition procedure. The traditional method for analyzing MRS signals uses the software package LCModel to estimate the absolute concentrations of these metabolites from the amplitudes and widths of the peaks in the spectrum. This approach assumes that the signal arises from a known set of metabolites and finds the best fit to a collection of pre-defined basis functions representing this set. LCmodel comparison between the different groups did not show any statistical differences. Our approach makes no assumptions about the underlying metabolite population, and instead extracts a rich set of wavelet-based features from the entire MRS signal, generating significantly more candidate biomarkers. By capturing the structure of all significant peaks in the signal, the wavelet-based method allows for the discovery of previously unknown signatures related to disease state that are not observed in the LCmodel group comparisons.

Conclusions:

MRS has been demonstrated to provide a non-invasive means of measuring brain biochemistry and by doing so provides a "virtual biopsy" to monitor a range of neurological diseases. In this study, we investigate the signatures for PTSD and mTBI, as compared to a robust set of controls using two different methods. Statistical analysis revealed significant group differences in the MRS signals across a wide range of metabolites. Compared to the LCModel approach, wavelet decomposition was able to identify significantly different regions of the spectra that can therefore be used for classification in future cohorts.

Acknowledgements:

This study was funded by DOD CDMRP WX81-XWH-10-1-0835. The views expressed in this abstract are those of the authors and do not reflect the official policy of the Department of Army, Department of Defense, or the U.S. Government.

APPENDIX F

"Differences in Brain Biochemistry between Military and Civilian Controls." AP Lin, H Liao, J Saurman, P Merugumala, I Orlovsky, X Long, S Merugumala, K Rudolph, B Rowland, K Heaton. Presented as a poster at the Military Health System Research Symposium (MHSRS). Florida, August 2015.

Abstract:

BACKGROUND: Many neuroimaging studies have examined traumatic brain injury and post-traumatic stress disorder in military subjects. Several research studies had used civilian subjects as a control group and found significant differences. However this may introduce bias to study results since there may be inherent differences between military and civilian subjects. To our knowledge, there has not been a study that challenges the assumptions that the cohorts are the same. Using magnetic resonance spectroscopy, a non-invasive method of brain chemistry, our aim was to determine if there are inherent biochemical differences between the two cohorts.

METHODS: 9 healthy military subjects (including service members and veterans, mean age 32.1±9.7, 3 female, 6 male) and 9 age- and gender-matched healthy civilian controls (mean age 32.7±11.6) were recruited and consented under local IRB approval. Both healthy military and civilian subjects had no history of neurological disorders, psychological disorders or drug addiction by self-report. All subjects underwent neuropsychological evaluation.

This study was performed in a 3T scanner and using 32-channel head coil. MRS was acquired (PRESS, TE=30ms, TR=2s, bandwidth=1.2kHz, 1024 complex data points, and 128 averages using 20x20x20 mm^3) in four different brain regions: Posterior Cingulate Gyrus (PCG), Posterior White Matter (PWM), Anterior Cingulate Gyrus and Left-temporal Lobe (20x15x15mm). Unsuppressed water spectrum with the same parameters but without water suppression and 16 averages. MRS data was analyzed using LCmodel. Metabolite concentrations were expressed in institutional units as well as a ratio of metabolite to total creatine (Cr+PCr).

RESULTS: Among all LC-model quantifiable metabolites, glutamate and NAA concentration showed significant differences between healthy military and civilian subjects. Compared to civilian subjects, lower Glu/Cr+PCr and NAA+NAAG/Cr+PCr ratios were observed in military subjects in all four voxel locations and significantly in PCG ($p<0.05$) and PWM ($p<0.001$). Glutamate and NAA were not highly correlated with each other ($R2=0.54$ in PWM). Cr+PCr was not found to be significantly different. All healthy civilian and military subjects were negative for post-concussive symptoms, PTSD, and depression. There were no significant differences between the two groups in their performance on neuropsychological testing.

CONCLUSION: Glutamate and NAA both had lower mean ratios in the military group across all four voxel locations and the most significant reduced mean ratios in PWM. They were not correlated therefore likely are independent measures. Given that depression and intelligence scores were not different, these changes cannot be attributed to those factors. As glutamate has been found to be increased in subjects with TBI and PTSD, it is possible that this reduction may reflect a cognitive reserve, possibly as a result of military training, providing physiological evidence of psychological resilience, though this remains to be proven. These results refute the assumption that the two groups are the same. Military studies should utilize healthy controls with similar military background.

ACKNOWLEDGEMENTS: This study was funded by DOD CDMRP WX81-XWH-10-1-0835. The views expressed in this abstract are those of the authors and do not reflect the official policy of the Department of Army, Department of Defense, or the U.S. Government.

APPENDIX G

Mariano L, Irvine J, Rowland B, Liao V, Ladner J, Heaton K, Lin A. "Novel Processing of Magnetic Resonance Spectroscopy Signal Enables Biomarker Discovery for PTSD and mTBI" Poster presented at AMSUS, 30 Nov - 1 Dec 2015, San Antonio, TX.

Abstract:
Post Traumatic Stress Disorder (PTSD) and mild Traumatic Brain Injury (mTBI) affect a large number of returning soldiers from Operation Iraqi Freedom and Enduring Freedom (OIF / OEF). Our study focuses on magnetic resonance spectroscopy (MRS) measurements to distinguish subjects having mTBI, PTSD, or both, with the goal of discovering biomarkers from the MRS data. MRS provides a non-invasive *in vivo* technique for measuring metabolite concentrations in the brain to aid in understanding neurological disorders.

Using a simple factorial design, subjects fall into five categories: military and civilian controls, military subjects diagnosed with PTSD only, mTBI only, and both PTSD and mTBI. Acquisition of the MRS data was performed on a Siemens Verio 3T scanner using a 32 channel head coil. Data were extracted in the Siemens 'twix' format which contains individual free induction decays for each channel and average. Channel weightings for the raw data were determined from the water reference signal using a singular value decomposition method designed to maximize SNR, then applied to the main data. Features were extracted from each signal using two approaches: wavelet decomposition, and LCModel. MRS features from subjects with PTSD, mTBI, or both were compared against data from control subjects.

MR spectroscopy signals of the brain are modeled as a superposition of the resonances from the underlying metabolites, plus distortions arising from the data acquisition procedure. The traditional method for analyzing MRS signals uses the software package LCModel to estimate the absolute concentrations of these metabolites from the amplitudes and widths of the peaks in the spectrum. This approach assumes that the signal arises from a known set of metabolites and finds the best fit to a collection of pre-defined basis functions representing this set. Using the LCModel approach, significant differences were found between the civilian and military controls but no statistical differences were found between the other groups. Healthy and neurologically normal military controls showed significantly lower glutamate, an excitotoxic neurotransmitter, compared to age, gender, and education matched civilian controls. There was also a weaker but significant reduction of N-acetyl aspartate, a neuronal marker, in the military cohort. No other differences were observed in the mTBI, PTSD, or mTBI+PTSD groups.

We developed a new approach for analyzing MRS signals that makes no assumptions about the underlying metabolite population, and instead extracts a rich set of wavelet-based features from the entire MRS signal, generating significantly more candidate biomarkers. By capturing the structure of all significant peaks in the signal, the wavelet-based method allows for the discovery of previously unknown signatures related to disease state that are not observed in the LCModel group comparisons. Our wavelet decomposition approach confirmed LCModel findings when comparing the two control cohorts but also identified significantly different regions of the spectra when comparing military controls to military subjects with PTSD. Similarly, significant differences were found between the PTSD group and the mTBI group. However, no significant difference was found between military controls and the mTBI group.

*Acknowledgements: T*his study was funded by DOD CDMRP WX81-XWH-10-1-0835. The views expressed in this abstract are those of the authors and do not reflect the official policy of the Department of Army, Department of Defense, or the U.S. Government.

APPENDIX H

Quatieri TF, Williamson JR, Smalt CJ, Helfer BS, Patel T, Perricone J, Ciccarelli G, Mehta DD, Ricke D, Malyska N, Palmer J, Heaton K, Eddy M, Moran J. "Comparison of vocal and EEG biomarkers to discriminate cognitive load in a working memory task." Poster presented at the Body-Sensor-Network, MIT Media Lab, Cambridge MA, 9-11 June 2015.

Comparison of Vocal and EEG Biomarkers to Discriminate Cognitive Load in a Working Memory Task*

Thomas F. Quatieri[1], James R. Williamson[1], Christopher J. Smalt[1], Brian S. Helfer[1], Tejash Patel[1], Joseph Perricone[1], Greg Ciccarelli[1], Daryush D. Mehta[1], Darrell Ricke[1], Nicolas Malyska[1], Jeff Palmer[1], Kristin Heaton[2], Marianna Eddy[3], Joseph Moran[3]
[1]MIT-Lincoln Laboratory, Lexington, MA, [2]USARIEM, [3]NSRDEC

Summary

In this study, a novel set of vocal biomarkers are introduced for detecting different cognitive load conditions. Our vocal biomarkers use phoneme- and pseudosyllable-based measures, and articulatory and source coordination derived from cross-correlation and temporal coherence of formant and creakiness measures. A ~2-hour protocol was designed to induce cognitive load by stressing auditory working memory. This was done by repeatedly requiring the subject to recall a sentence while holding a number of digits in memory. We demonstrate the power of our speech features to discriminate between high and low load conditions. Using a database consisting of audio from 13 subjects, we apply classification models of cognitive load, showing a 7% detection equal-error rate from features derived from 40 sentence utterances (~4 minutes of audio).

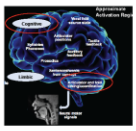
Introduction

Cognitive load and stress

- Cognitive load is defined loosely as the mental demand experienced for a particular task
- More efficient and effective methods are needed to monitor cognitive load under cognitively and physically stressful conditions.
- Such conditions include environmental and occupational stressors that can result in dangerous scenarios when under heavy cognitive load

Objectives

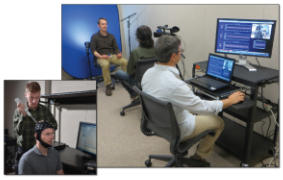
- Detect different cognitive load conditions from speech/voice, EEG, face, physiology
- Predict cognitive fatigue from these modalities



Data and Protocol

Database

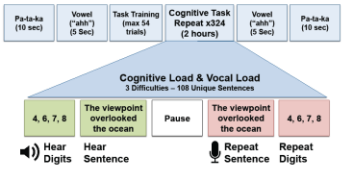
- Goal of 20 subjects at MIT LL
- 13 subjects collected thus far



Cognitive task involves auditory working memory:

1. Hear digits
2. Hear sentence
3. Recall sentence
4. Recall digit

Three levels of difficulty invoke subject-dependent digit spans from 2-7 digits



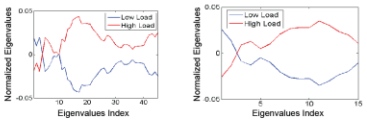
Feature Extraction and Correlation Analysis

Vocal analysis: During sentence recall

- Phonetic timing
- Articulatory coordination and coherence (resonances)
- Articulatory (resonances) – source (creak) coordination

Correlation structure feature examples

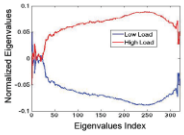
Average normalized eigenvalues from all subjects for low and high cognitive loads, based on articulatory coordination (left) and articulatory-source coordination (right).



EEG analysis: During pause analysis

- Correlation structure
- Coherence structure
- Log-power

Average normalized eigenvalues from all subjects for low and high cognitive loads



- Avoids motion and muscle artifacts
- 64 channel Neuroscan Unit

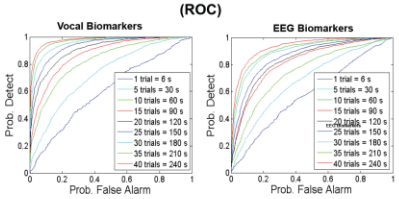
Classification of Cognitive Load

- To evaluate detection performance, for each subject:
  - The 108 feature vectors (one vector per spoken sentence and load condition) from the max-digit condition is assigned to the high load class
  - The 108 vectors from the max-digit-minus-two condition is assigned to the low load class.
- Leave-one-subject-out cross-validation is used with a classifier trained on the data from 12 held out subjects when assessing ability to discriminate between high and low load on a test subject
- Load discrimination is done with a Gaussian classifier (GC), where the Gaussians are centered on the two class means, and a common covariance matrix is used based on the data across both load conditions

*This work is sponsored by the Assistant Secretary of Defense for Research & Engineering under Air Force contract #FA8721-05-C-0002. Opinions, interpretations, conclusions, and recommendations are those of the authors and are not necessarily endorsed by the United States Government.

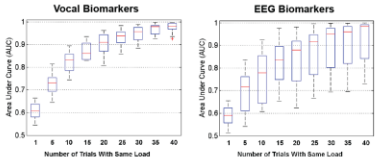
Cognitive Load Detection

Performance Evaluation: True Detection vs. False Alarm (ROC)



Results as a function of number of combined trials with same load.

Performance Evaluation: Area under the ROC curve



Results as a function of number of combined trials with same load.

Conclusions and Future Work

- Vocal timing and correlation/coherence and EEG correlation/ coherence features can detect high and low load conditions after about 4 minutes of repeated trials
- For this small pilot, vocal features outperform EEG in terms at small time intervals
- Larger data sets required for significance
- Expansion to physiologic and facial features planned

References

1. Lively, S.E., Pisoni, D.B., Van Summers, W., Bernacki, R.H., "Effects of cognitive workload on speech production: Acoustic analyses and perceptual consequences," J Acoust Soc Am. 1993 May; 93(5): 2962–2973.
2. Yin, B., Chen, F., Ruiz, N., Ambikairajah, E., "Speech-based cognitive load monitoring system, ICASSP 2008.
3. Khawaja, M.A., Ruiz, N., Cheng, F., "Think before you talk: An empirical study of relationship between speech pauses and cognitive load," OZCHI 2008, December 8-12, 2008.
4. Le, P., J. Epps, Choi, H.C., and Ambikairajah, E., "A study of voice source- and vocal tract-based features in cognitive load classification," Proceedings of International Conference on Pattern Recognition, 2010, pp. 4516-4519.
5. Trevino, A., Quatieri, T. F. and Malyska, N., "Phonologically-based biomarkers for major depressive disorder," EURASIP Journal on Advances in Signal Processing: Special Issue on Emotion and Mental State Recognition from Speech, 42:2011–2042, 2011
6. Williamson, J. R., Quatieri, T. F., Helfer, B. S., Ciccarelli, G., & Mehta, D. D. (2014, November). Vocal and Facial Biomarkers of Depression based on Motor Incoordination and Timing. In Proceedings of the 4th International Workshop on Audio/Visual Emotion Challenge (AVEC), pp. 65-72. ACM
7. Williamson, J.R., Bliss, D., Browne, D.W., and Narayanan, J.T., "Seizure prediction using EEG spatiotemporal correlation structure," Epilepsy Behav., vol. 25, no. 2, pp. 230–238, 2012.
8. Kane, J., Drugman, T., & Gobl, C. "Improved automatic detection of creak," Computer Speech & Language, 27:1028-1047, 2013.

APPENDIX I

Vocal biomarkers to discriminate cognitive load in a working memory task

Thomas F. Quatieri [1], James R. Williamson[1], Christopher J. Smalt[1], Tejash Patel[1], Joseph Perricone[1], Daryush D. Mehta[1], Brian S. Helfer[1], Greg Ciccarelli[1], Darrell Ricke[1], Nicolas Malyska[1], Jeff Palmer[1], Kristin Heaton[2], Marianna Eddy[3], Joseph Moran[3]

[1]MIT Lincoln Laboratory, Lexington, Massachusetts, USA
[2]USARIEM, [3]NSRDEC

[quatieri,jrw]@ll.mit.edu

Abstract

Early, accurate detection of cognitive load can help reduce risk of accidents and injuries, and inform intervention and rehabilitation in recovery. Thus, simple noninvasive biomarkers are desired for determining cognitive load under cognitively complex tasks. In this study, a novel set of vocal biomarkers are introduced for detecting different cognitive load conditions. Our vocal biomarkers use phoneme- and pseudosyllable-based measures, and articulatory and source coordination derived from cross-correlation and temporal coherence of formant and creakiness measures. A ~2-hour protocol was designed to induce cognitive load by stressing auditory working memory. This was done by repeatedly requiring the subject to recall a sentence while holding a number of digits in memory. We demonstrate the power of our speech features to discriminate between high and low load conditions. Using a database consisting of audio from 13 subjects, we apply classification models of cognitive load, showing a 7% detection equal-error rate from features derived from 40 sentence utterances (~4 minutes of audio).

Index Terms: cognitive load, vocal biomarkers, phoneme and pause duration, articulatory coordination

1. Introduction

Cognitive load is defined loosely as the mental demand experienced for a particular task [1][2]. More efficient and effective methods are needed to monitor cognitive load under cognitively and physically stressful conditions. Such conditions include environmental and occupational stressors that can result in dangerous scenarios when cognitively overloaded. Examples of mental stressors are repetitive and/or intense cognitive tasks, psychological stress, and lack of sleep. Physical stressors include intense long-duration operations and/or heavy loads. Both stressors can cause cognitive load, and often contribute simultaneously to load. Applications for cognitive load assessment include individualized detection of cognitive load in an ambulatory, field, or clinical setting. In clinical applications, the objective is often to find and measure the specific causes of load. In operational settings, the objective is often to quickly assess cognitive ability and readiness under loaded conditions, regardless of their etiology.

Biomarkers for monitoring and detecting cognitive load comprise behavioral, physiologic, and cognitive modalities. A potential class of biomarkers that has recently gained popularity is based on speech characteristics. Vocal features are desirable as biomarkers of cognitive status because they can be obtained easily (e.g., via telephone), greatly increasing global accessibility to an automated method for cognitive assessment. Certain vocal features have been shown to change with a subject's mental and emotional state, under numerous conditions including cognitive load. These features include characterizations of prosody (e.g., fundamental frequency and speaking rate), spectral representations (e.g., mel-cepstra), and glottal excitation flow patterns, such as flow shape, timing jitter, amplitude shimmer, and aspiration [1]-[8].

A motivation for the vocal features developed in the present paper is the hypothesis that cognitive load can be assessed by measures of speech-segment-based prosodic dynamics and articulatory and source coordination. Specifically, we employ phoneme- and pseudosyllable-based measures that include rate, duration and pitch dynamics, as well as pause information, and articulatory and source coordination measures from cross correlations and cross coherences among extracted signals such as formant tracks, delta mel-cepstra coefficients, and creakiness signals. A subset of these vocal features have been used effectively in other neuro-cognitive contexts such as in detection of depression, traumatic brain injury, and dementia [9]-[11], thus perhaps forming a common vocal feature basis for neurocognitive change.

Our paper is organized as follows. In Section 2, we describe our data collection using a novel cognitive load protocol that taxes auditory working memory by eliciting sentence recall under varying levels of cognitive load. In Section 3, we describe our signal processing methodologies for vocal feature extraction. Section 4 reports cognitive load detection results using a Gaussian classifier. Section 5 provides conclusions and projections toward future work.

2. Working Memory Protocol

Subjects gave informed consent to a working memory-based protocol approved by the MIT Committee on the Use of Humans as Experimental Subjects (COUHES). Audio data are collected with a DPA acoustic lapel microphone (with a Roland Octa-Capture audio

*This work is sponsored by the Assistant Secretary of Defense for Research & Engineering under Air Force contract #FA8721-05-C-0002. Opinions, interpretations, conclusions, and recommendations are those of the authors and are not necessarily endorsed by the United States Government.

interface)*. Following setup and training, each subject engages in the primary task of verbally recalling sentences with varying levels of cognitive load, as determined by the number of digits being held in working memory [18]. Specifically, a single trial of the auditory working memory task comprises: the subject hearing a string of digits, then hearing a sentence, then waiting for a tone eliciting spoken recall of the sentence, followed by another tone eliciting recall of the digits. This task is administered with three difficulty levels, involving 108 trials per level. The same set of 108 sentences is used in each difficulty level. The order of trials (sentences and difficulty level) is randomized. The entire protocol, approximately two hours in duration, is illustrated in Figure 1. The multi-talker PRESTO sentence database is used for sentence stimuli [15].

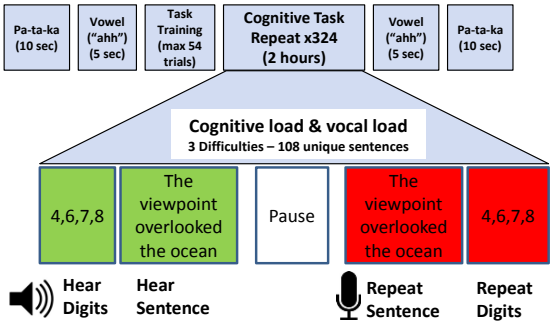


Figure 1. Auditory working memory protocol.

The working memory task is split into a training and a testing phase. During training, the length of presented digit sequence is alternately increased and decreased via an adaptive tracking algorithm [16] to determine the geometric mean of the number of digits a subject can recall. This number, n_c, is used to determine the maximum number of digits presented during testing, d_n, used in the three difficulty levels for the test phase. The test phase consists of 324 (= 108x3) consecutive trials randomized and balanced across the three difficulty levels.

We used $d_n = \{\text{floor}(n_c), \text{floor}(n_c)\text{-}2, \text{floor}(n_c)\text{-}4\}$ for the first four subjects. Thus, a subject with n_c=5.32 would have d_n={5, 3, 1}. One way these difficulty levels might manifest behaviorally is via subject accuracy, with poorer accuracy at higher loads (more digits). The performance differences between easy and difficult conditions were small for our subjects, so we increased the difficulty by using $d_n = \{\text{ceil}(n_c), \text{ceil}(n_c)\text{-}2, \text{ceil}(n_c)\text{-}4\}$. This modification was used for one subject, after which we adopted $d_n = \{\text{ceil}(n_c), \text{ceil}(n_c)\text{-}1, \text{ceil}(n_c)\text{-}2\}$ for the remaining eight subjects. As a result of these changes, digit-span accuracy was consistently lower for the hardest difficulty level compared to the easiest among the final nine subjects.

*The audio collection is part of a larger multi-modal collection involving also facial video, EEG, and physiologic modalities with analysis on-going.

Despite the minor protocol changes between early and late subjects, a common load assessment test for all 13 subjects is possible due to the fact that all subjects had both a *max number* condition and a *max number minus two* condition. The range of digit spans across all subjects was 2–5 for low load and 4–7 for high load.

3. Feature extraction

Feature vectors are extracted only from the single spoken sentence component of each trial in the test phase of the auditory memory task. *Low-level* vocal features comprise measures of phoneme durations, pseudosyllable rate, pitch dynamics, articulation, spectral dynamics, and creak. We construct *high-level* features that capture inter-relationships among the low-level features. The feature sets are derived under the hypothesis that differences in cognitive load produce detectable changes in speech production timing and articulatory and source coordination. Low-level features, produced every 10 ms, are approximately immune to slowly-varying linear channel effects due to not being directly dependent on spectral magnitude.

3.1 Low-level vocal feature extraction

Phonemes: Using an automatic phoneme recognition algorithm [12], phonetic boundaries are detected, with each segment labeled with one of 40 phonetic speech classes.

Pseudo-syllables: Vocal syllable-like patterns are detected based on the concept of a pseudo-syllable (PS) [19]. The automatic phoneme recognition system detects individual speech sounds, which are combined into PS segments. For example, "v," "cv," and "ccv" are all valid PSs.

Pitch slopes: The fundamental frequency (pitch) is estimated using an autocorrelation method over a 40-ms Hanning window every 1 ms [20]. Within each phone or PS segment, a linear fit is made to the log of the pitch, yielding a pitch slope (Δlog(Hz)/s) for each phonetic or PS speech unit.

Formant frequencies: A Kalman filter technique is used to characterize vocal tract resonance dynamics by smoothly tracking the first three formant frequencies, while also smoothly coasting through non-speech regions [13].

Mel-frequency cepstral coefficients (MFCCs): 16 delta MFCCs are used to characterize velocities of vocal tract spectral magnitudes. Delta MFCCs are computed using regression with the two frames before and after a given frame.

Creak voice quality: A creaky voice quality (vocal fry, irregular pitch periods, glottalization, etc.), is characterized using acoustic measures of low-frequency/damped glottal pulses [21]. Low-level features include previously-studied metrics of short-term power, intra-frame periodicity, inter-pulse similarity [22], and two measures of the degree of sub-harmonic energy (reflecting the presence of secondary glottal pulses) and the temporal peakiness of glottal pulses with long period [23]. These low-level features are input into

an artificial neural network to yield creak posterior probabilities on a frame-by-frame basis [24].

3.2 High-level features

Our high-level features are designed to characterize properties of timing and coordination from the low-level features.

Phoneme-dependent features: Building on previous work [8]-[11], features conditioned on time segments of detected phonemes are constructed based on their discriminative value. For each phoneme, the features considered are: phoneme counts, total phoneme durations, and slopes of log-pitch during phonemes [9][11]. These features were computed in two different conditions: 1) for all detected phonemes, and 2) for those phoneme instances where pitch slopes are marked as valid. Based on [9], the slope of log pitch values is marked as valid if its absolute value is less than eight, indicating that the slope is likely derived from a continuous pitch contour.

Four phoneme-based features were found useful, each an aggregate derived from a linear combination of 25 phonemes, with weights based on their discriminative value. In [9] these weights were derived from correlations with depression scores. Here, each weight is the signed Mahalanobis distance between the measured distributions (using mean and variance) for high and low loads. Table 1 lists the five most important phonemes and their weights for each of the aggregate features.

Table 1. Phoneme-based features. The top 5 phonemes are listed for each feature, along with their weights.

All Phns		Phns with valid pitch slopes					
Phn count		Phn count		Phn dur.		Pitch slope	
Phn	*w*	Phn	*w*	Phn	*w*	Phn	*w*
‘sil’	2.2	‘v’	1.5	‘v’	1.4	‘ae’	1.1
‘v’	1.4	‘ch’	1.1	‘ch’	1.2	‘ay’	1.1
‘hh’	-1.2	‘w’	-1.0	‘w’	-1.0	‘ng’	1.0
‘zh’	0.8	‘zh’	1.0	‘zh’	0.9	‘d’	0.8
‘sh’	-0.7	‘hh’	-1.0	‘ao’	-0.8	‘k’	0.7

It is interesting to observe that the total pause count (“sil”) plays an important role, consistent with other findings [4].

Pseudosyllable-based features: A similar processing approach is applied to pseudosyllable (PS) speech segments. The PS dictionary contains silence (‘#’) and different combinations of consonants (‘c’) and vowels (‘v’). Three different aggregate PS features were found useful, based on linear combinations of the top 10 PS-based measures of counts and pitch dynamics. As with the phoneme-based features, weights are the signed Mahalanobis distances between the measures for high and low loads (Table 2). Again the total pause count (“sil”) plays an important role.

Table 2. Pseudosyllable (PS)-based features. For each feature, the top five PS-based measures are listed, along with their weights.

All PS		PS with valid pitch slopes			
PS count		PS count		PS pitch slope	
PS	*w*	PS	*w*	PS	*w*
‘#’	2.2	‘c’	1.7	‘ccv’	0.7
‘c’	1.5	‘ccv’	-0.8	‘cccv’	-0.6
‘ccc’	1.0	‘ccc’	0.7	‘v’	0.5
‘ccccccv’	-0.7	‘cccccccv’	-0.7	‘cccc’	-0.5
‘ccv’	-0.7	‘ccccccv’	0.7	‘ccc’	-0.4

Correlation Structure: Measures of the structure of correlations among low-level speech features have been applied in the estimation of depression [9], the estimation of cognitive performance associated with dementia [8], and the detection of changes in cognitive performance associated with mild traumatic brain injury [10]. The details for this approach are in [25], where the method was first introduced for analysis of EEG signals for epileptic seizure prediction.

Channel-delay correlation and covariance matrices are computed from multiple time series channels of vocal parameters. Each matrix contains correlation or covariance coefficients between the channels at multiple time delays. Changes over time in the coupling strengths among the channel signals cause changes in the eigenvalue spectra of the channel-delay matrices. The matrices are computed at multiple “time scales” corresponding to separate sub-frame spacings. Features at each time scale consist of the eigenvalue spectra of channel-delay *correlation* matrices, as well as covariance power (logarithm of the trace) and entropy (logarithm of the determinant) from channel-delay *covariance* matrices.

In previous applications, vectors comprising the correlation-based eigenspectra and covariance-based entropy and power have been concatenated into a single feature vector and then projected, using principal component analysis (PCA), into a lower-dimensional representation. In the current application, better discriminative value was found by applying PCA separately to the multi-scale correlation-based and covariance-based features.

Table 3 shows parameters used to extract correlation structure features from three different low-level speech sources: formant frequency tracks, creak probabilities, and delta MFCCs. Sub-frame spacings of 1, 3, 7, 15, and 21 are used and, due to the 10-ms frame interval of the low-level features, these correspond to time spacings of 10, 30, 70, 150, and 210 ms, respectively. Each matrix (for each scale) is constructed using 15 time delays. The number of correlation-based features is the number of signal channels times the number of scales (i.e., number of sub-frame spacings) times the number of time delays (15) per time scale. The number of covariance-based features is the number of time scales

(entropy features) plus one log power feature, as power is invariant across scale. Parameters are similar to those of previous studies [8]-[10], with numbers of principal components empirically chosen based on discrimination performance.

The differences in eigenspectra patterns due to high versus low cognitive loads provide indications about the effect of load on speech. In Figure 2, averages across all subjects of normalized eigenvalues from formant and creak signals at time scale 3 (sub-frame spacing of 7) are shown for low load (blue) and high load (red). The eigenvalues are ordered, left to right, from largest to smallest. So, in both cases, there is greater power in the small eigenvalues during higher cognitive load. This indicates greater dynamical complexity in formant frequencies and creak during higher cognitive load. The y-axes are in units of standard deviation.

Table 3. Channel-delay correlation and covariance features.

Signal type	# chan-nels	Feat. type	Sub-frame spacings	# raw feat.	# PCA feat.
Formant	3	corr.	1,3,7	135	3
Creak	1	corr.	1,3,7,15,31	75	2
Fmt-Crk	4	corr.	1,3,7	180	3
Formant	3	cov.	1,3,7	4	3
dMFCC	16	cov.	1,3,7	4	4

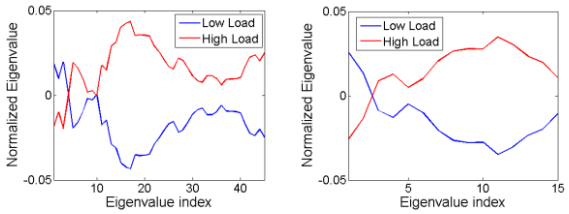


Figure 2. Correlation structure features: Average normalized eigenvalues from all subjects for low and high cognitive loads, based on formant frequencies (left) and creak (right).

Coherence Structure and Power:
We have also introduced a feature set that characterizes the structure of signal coherence and power at multiple frequency bands. The coherence between channels, indicating the amount of cross-channel power in a frequency band relative to the amount of within-channel power, provides a measure of how closely related the signals are within a frequency band. The power and cross-power are computed among three formant frequency channels in two different frequency bands, and a 3 × 3 coherence matrix is constructed for each band. The eigenspectra of the coherence matrices indicate the structure of coherence across the channels. PCA is used to project these features into lower dimensional representations. Table 4 indicates the parameters, selected empirically by performance measures, used for the coherence and power features.

The differences in coherence and power features due to high versus low cognitive load provide indications about the effect
of load on speech. In Figure 3 (left), averages across all subjects of normalized coherence eigenvalues from frequency band 1.0–2.0 Hz are shown for low load (blue) and high load (red). The eigenvalues are ordered, left to right, from largest
to smallest. Similar to the correlation structure results shown in Figure 2, these results indicate greater power in the smaller

Table 4. Frequency band coherence and power features.

Signal type	Feature type	Freq. Band (Hz)	# raw features	# PCA features
Formant	Coh.	0.25 – 1.0	3	1
Formant	Coh.	1.0 – 2.0	3	1
Formant	Log Pow.	1.0 – 2.0	3	2

eigenvalues for the higher load condition. In Figure 3 (right), it is shown that the higher load condition is also associated with less power for the first two formants. The y-axes are in units of standard deviation.

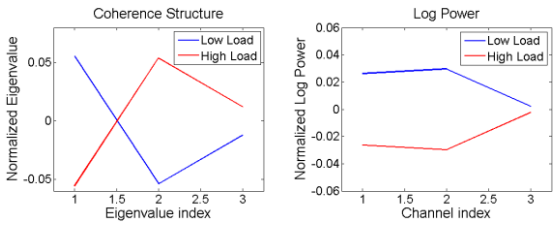


Figure 3. Left: Average normalized eigenvalues from coherence matrix at frequency band 1.0–2.0 Hz for low and high cognitive loads from formant frequencies. Right: Normalized log power for the three formant frequencies at frequency band 1.0–2.0 Hz.

4. Results

Our goal is to detect differences in cognitive load from voice measurements. To evaluate detection performance, for each subject the 108 feature vectors (one vector per spoken sentence and load condition) from the max-digit condition is assigned to the *high* load class, and the 108 vectors from the max-digit-minus-two condition is assigned to the *low* load class. Leave-one-subject-out cross-validation, is used, with a classifier trained on the data from 12 held out subjects when

assessing ability to discriminate between high and low load on a test subject.

A key processing step is individualized feature normalization. This involves, for each subject (whether in the training or test set), subtracting the mean from each feature across both load conditions. This processing step is done to remove inter-subject feature variability, and implies that the ability to discriminate load conditions requires some knowledge of a subject's baseline features.

Load discrimination is done with a Gaussian classifier (GC), where the Gaussians are centered on the two class means, and a common covariance matrix is used based on the data across both load conditions. In each trial, the GC produces a load score (log-likelihood ratio of high versus low load). A receiver operating characteristic (ROC) curve is obtained by varying a detection threshold to characterize the sensitivity/specificity tradeoff. For each subject, 216 scores are obtained (108 for each load). A single ROC curve derived from scores of all 13 subjects characterizes total performance, with the area under the curve (AUC) serving as a summary statistic.

In Table 5 is listed the number of features used by the GC for each feature set, and the AUC results. The feature sets consists of the features described in Tables 1-4. The best overall performance of AUC = 0.61 is obtained by combining (via vector concatenation) all four feature sets.

Table 5. Summary of area under ROC curve (AUC) results for detecting high cognitive load from a single trial (sentence).

Feature Set	# features	AUC
Phoneme-based	4	0.59
PS-based	3	0.55
Corr. structure	15	0.56
Coh. structure	4	0.54
Combined	**26**	**0.61**

Although our protocol involves feature processing of single spoken sentences, the ability to detect load across multiple sentences can be assessed by combining the GC scores from different trials, provided that the trials involve the same load condition. This was done by randomly selecting, from the same subject, a number of trials of either high load or low load, and summing their GC scores. For each subject, load condition and combination number, 200 randomly chosen sets of trials were used to determine performance across multiple sentences. Figure 4 (left) contains boxplots summarizing the AUC values for the 13 subjects, given combinations of 1, 5, 10, ..., 40 trials. The median AUC value is 0.83 after 10 trials and 0.91 after 20 trials, with AUC for all subjects > 0.9 after 35 trials. In Figure 4 (right) are shown the cross-subject ROC curves from the same multi-trial combinations. For 40 trials (~4 minutes), we obtain an equal error rate of ~7%, corresponding to ~93% detection with ~7% false alarm.

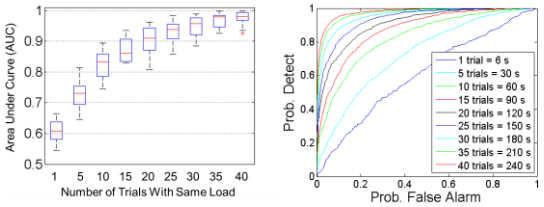


Figure 4. Results as a function of number of combined trials with same load. Left: AUC values across 13 subjects. Right: cross-subject ROC curves.

5. Conclusions and Discussion

In this paper, we demonstrated the power of our speech features to discriminate between high and low load conditions. Our features capture inter-relationships among phoneme durations, pseudosyllable rates, pitch dynamics, articulation, spectral dynamics, and creak. Using a database consisting of audio from 13 subjects and recalled sentences prior to recalling a digit span, we effectively applied classification models of cognitive load. For example, with 40 trials (~4 minutes), we obtain an equal error rate of ~7%, corresponding to ~93% detection with ~7% false alarm.

As mentioned in the Introduction, there has been prior work in use of vocal features in detecting cognitive load [1]-[7]. For example, Yin et al [2] achieved 77% accuracy discriminating 3 cognitive load levels over a read story and several questions about the story and over the Stroop test using standard vocal features (e.g., mel-cepstra, delta-delta mel-cepstra, and shifted mel-cepstra). Our approach, on the other hand, uses standard features at a "low-level" from which relational information is derived. Future work will involve a more formal comparison with alternative conventional approaches. Future work will also involve expansion of our approach to other modalities that form our larger data collection (EEG, facial video, and physiology).

6. References

1. Lively, S.E., Pisoni, D.B., Van Summers, W., Bernacki, R.H., "Effects of cognitive workload on speech production: Acoustic analyses and perceptual consequences," J Acoust Soc Am. 1993 May; 93(5): 2962–2973.
2. Yin, B., Chen, F., Ruiz, N., Ambikairajah, E., "Speech-based cognitive load monitoring system, ICASSP 2008.
3. Yin, Bo, and Fang Chen. "Towards automatic cognitive load measurement from speech analysis." *Human-Computer Interaction. Interaction Design and Usability*. Springer Berlin Heidelberg, 2007. 1011-1020.
4. Khawaja, M.A., Ruiz, N., Cheng, F., "Think before you talk: An empirical study of relationship

between speech pauses and cognitive load," OZCHI 2008, December 8-12, 2008.
5. Le, P., J. Epps, Choi, H.C., and Ambikairajah, E., "A study of voice source- and vocal tract-based features in cognitive load classification," Proceedings of International Conference on Pattern Recognition, 2010, pp. 4516-4519.
6. Boril, H., Sadjadi, O., Kleinschmidt, T., and J. Hansen, Analysis and detection of cognitive load and frustration in drivers' speech," Proceedings of Interspeech, 2010, pp. 502-505.
7. Yap, T.F., Speech Production Under Cognitive Load: E˙ects and Classi˝cation, PhD Thesis, The University of New South Wales School of Electrical Engineering and Telecommunications Sydney, Australia, Sept. 2011.
8. Yu, B., Quatieri, T.F., Williamson, J.W., and Mundt, J., "Prediction of cognitive performance in an animal fluency task based on rate and articulatory markers." Fifteenth Annual Conference of the International Speech Communication Association. 2014.
9. Williamson, J. R., Quatieri, T. F., Helfer, B. S., Ciccarelli, G., & Mehta, D. D. (2014, November). Vocal and Facial Biomarkers of Depression based on Motor Incoordination and Timing. In *Proceedings of the 4th International Workshop on Audio/Visual Emotion Challenge* (AVEC), pp. 65-72. ACM. (Winning paper in (AVEC) Depression Challenge)
10. Helfer, B. S., Quatieri, T. F., Williamson, J. R., Keyes, L., Evans, B., Greene, W. N., Palmer, J., & Heaton, K. (2014). Articulatory Dynamics and Coordination in Classifying Cognitive Change with Preclinical mTBI. In *Fifteenth Annual Conference of the International Speech Communication Association.*
11. Trevino, A., Quatieri, T. F. and Malyska, N., "Phonologically-based biomarkers for major depressive disorder," EURASIP Journal on Advances in Signal Processing: Special Issue on Emotion and Mental State Recognition from Speech, 42:2011–2042, 2011
12. Shen, W., White, C., Hazen, T.J., "A comparison of query-by-example methods for spoken term detection," in Proceedings of the 2010 IEEE International Conference on Acoustics Speech and Signal Processing (2010)
13. Mehta, D. D., Rudoy, D. and Wolfe, P. J., "Kalman-based autoregressive moving average modeling and inference for formant and antiformant tracking," The Journal of the Acoustical Society of America, 132(3):1732–1746, 2012.
14. Singer, J.D. and Willett, J.B., "Applied longitudinal data analysis: Modeling change and event occurrence," Oxford university press, 2003.
15. Park et al., 2010 H. Park, R. Felty, K. Lormore, D. Pisoni PRESTO: perceptually robust English sentence test: open set—design, philosophy, and preliminary findings J. Acoust. Soc. Am., 127 (2010), p. 1958
16. Levitt, H., 1971. Transformed up-down methods in psychoacoustics. J. Acoust. Soc. Am. 49, 467–477.
17. Le, P.N., Ambikairajah, E, Choi, H.C., and J. Epps, "A non-uniform sub-band approach to speech-based cognitive load classification," Proceedings of ICICS, 2009, pp. 1-5.
18. Harnsberger, James D., Richard Wright, and David B. Pisoni. "A new method for eliciting three speaking styles in the laboratory." *Speech communication* 50.4 (2008): 323-336.
19. Rouas J., "Automatic Prosodic Variations Modeling for Language and Dialect Discrimination, IEEE Trans.Audio, Speech, and Language Proc., Vol. 15, Nop. 6, August 2007.
20. Boersma, P. 1993. Accurate short-term analysis of the fundamental frequency and the harmonics-to-noise ratio of a sampled sound. Proceedings of the Institute of Phonetic Sciences. 17, (1993), 97–110.
21. Gerratt, B. R., and Kreiman, J., "Toward a taxonomy of nonmodal phonation," *Journal of Phonetics*, vol. 29, no. 4, pp. 365-381, 2001.
22. Ishi, C.T., Sakakibara, K.I., Ishiguro, H., and Hagita, N., "A method for automatic detection of vocal fry," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 16, no. 1, pp. 47-56, 2008.
23. Kane, J., Drugman, T., and Gobl, C., "Improved automatic detection of creak," *Computer Speech & Language*, vol. 27, no. 4, pp. 1028-1047, 2013.
24. http://tcts.fpms.ac.be/~drugman/Toolbox/.
25. Williamson, J.R., Bliss, D., Browne, D.W., and Narayanan, J.T., "Seizure prediction using EEG spatiotemporal correlation structure," *Epilepsy Behav.*, vol. 25, no. 2, pp. 230–238, 2012.

APPENDIX J

Horwitz-Martin R, Quatieri T, Helfer B, Williamson J, Vian T, Lacirignola J, Shenk T, Talavage T, Palmer J, Heaton K. "Phone durations as predictors of preclinical; mild traumatic brain injury symptom severity." Poster presented at the 5th Annual Traumatic Brain Injury Conference, Washington, DC, April 15-16, 2015.

Background

Injury sustained during athletic participation has become a major cause for concern at both the professional and sub-professional levels. As a result, current studies have sought to find early indicators of mild Traumatic Brain Injury (mTBI). Ongoing work at Purdue University with high school football players has demonstrated that cumulative sub-concussive impacts are associated with neurocognitive and neurophysiological impairment [1]. Previous work at MIT Lincoln Laboratory (MIT LL) used the speech collected as part of the Purdue database to identify biomarkers of cognitive decline as measured by the Immediate Post-Concussion Assessment for Cognitive Testing (ImPACT) suite [2]. Features describing articulatory dynamics and precision were used to identify changes in visual motor speed, visual memory, and reaction time [3]. This work showed high fidelity in predicting cognitive change; however, it is possible to gain greater insight into TBI and its impacted areas by adding speech features whose origins are localized to other brain regions. The current work extends upon prior clinical and automatic classification research, which has demonstrated a relationship between head trauma and signals associated with speech production [4][5].

Methods

Under an Institutional Review Board (IRB) approved protocol, we examine data from a group of 35 high school athletes participating in football and soccer. The data is collected before the athletic season commences to provide a baseline measurement, and then additional data is collected throughout the season. This is performed as part of a platform designed at MIT Lincoln Laboratory [6]. The data investigated in this study include the athletes' phone durations while reading a standard passage, as well as the athletes' cognitive scores as measured by the ImPACT test. The ImPACT cognitive score modalities are following: verbal memory, visual memory, visual motor speed, and reaction time. In this study, features reflecting the change from baseline phone duration are extracted. The features are then combined based on their correlation with each of the cognitive modalities, and then incorporated into Gaussian classifiers to predict cognitive decline. Classification performance is then analyzed using receiver operating characteristic (ROC) curves through detection versus false alarm statistics.

Results

For the ROC curve, a detection (true positive) is defined as a correct prediction by the classifier that an athlete has declined in cognitive performance from his/her baseline score. Likewise, a false alarm is defined to occur when decline is erroneously predicted. Additionally, the area under the ROC curve (AUC) gives an overall sense of performance with AUC = 1.0 being ideal. Our computed ROC curves demonstrate high fidelity prediction of cognitive change using vocal phonetic timing features for the four components of ImPACT that were studied. The highest AUCs achieved are 0.895, 0.806, 0.939, and 0.900 for verbal memory, visual memory, visual motor speed, and reaction time scores respectively.

Conclusion

Our current study uses phonetic timing features to detect changes in cognitive performance. While no single phone most accurately predicts cognitive decline, the combination of two or three delta phones is able to predict cognitive change with high fidelity. Detecting changes in cognitive status through such non-invasive monitoring has potential benefit, in that cognitive changes that reflect accumulative damage are expeditiously measured without laborious cognitive testing, and providing features that indicate preclinical TBI, i.e., increased susceptibility to mTBI. Furthermore, the high detection rate of the classifier, while maintaining low false alarm rate, suggests it could be used as a screening tool to determine readiness to return to play, thereby

decreasing the athletes' risk for subsequent injury.

References

[1] T. M. Talavage, E. A. Nauman, E. L. Breedlove, U. Yoruk, A. E. Dye, K. Morigaki, H. Feuer, and L. J. Leverenz, "Functionally-Detected Cognitive Impairment in High School Football Players Without Clinically- Diagnosed Concussion," Journal of Neurotrauma, no. 765, pp. 1–46, 2010.
[2] I. A. Inc, "Immediate Post-Concussion Assessment Testing (ImPACT ®) Test Technical Manual Technical Manual and Psychometric Data," impacttest.com, 2013.
[3] B.S. Helfer, T.F. Quatieri, J.R. Williamson, L. Keyes, B. Evans, W. N. Greene, T. Vian, J. Lacirignola, T. Shenk, T.Talavage, J. Palmer, K. Heaton. "Articulatory Dynamics and Coordination in Classifying Cognitive change with Preclinical mTBI, " in Interspeech 2014.
[4] D. G. Theodoros, Traumatic brain injury: Associated speech, language, and swallowing disorders. Cengage Learning, 2001.
[5] M. Falcone, N. Yadav, C. Poellabauer, and P. Flynn, "Using Isolated Vowel Sounds for Classification of Mild Traumatic Brain Injury," in Acoustics, Speech and Signal Processing (ICASSP), 2013 IEEE International Conference on, 2013.
[6] L. Keyes, J. Su, T. Quatieri, B. Evans, J. Lacirignola, T. Vian, W. Greene, D. Strom, and A. Dai, "FY12 Line-Supported Bio-Medical Initiative Program: Multi-modal Early Detection Interactive Classifier (MEDIC) for Mild Traumatic Brain Injury (mTBI) Triage," MIT Lincoln Laboratory Project

Disclaimer: This work is sponsored by the Assistant Secretary of Defense for Research & Engineering under Air Force contract #FA8721-05-C-0002. Opinions, interpretations, conclusions, and recommendations are those of the authors and are not necessarily endorsed by the United States Government.

APPENDIX K

ORIGINAL ARTICLES

Authors alone are responsible for opinions expressed in the contribution and for its clearance through their federal health agency, if required.

MILITARY MEDICINE, 180, 6:660, 2015

Neurocognitive Performance and Prior Injury Among U.S. Department of Defense Military Personnel

Susan P. Proctor, DSc†‡; Kenneth Nieto, MS*; Kristin J. Heaton, PhD*†; Caitlin C. Dillon*; Robert E. Schlegel, PhD§; Michael L. Russell, PhD§‖ ; Andrea S. Vincent, PhD¶*

ABSTRACT This study examined the neurocognitive performance of U.S. military personnel completing the Auto- mated Neuropsychological Assessment Metrics (version 4) TBI Military (ANAM4 TBI-MIL) battery as part of the Department of Defense Neurocognitive Functional Assessment Program. Descriptive analyses utilizing the ANAM4TBI Military Performance Database were performed. We examined ANAM Composite Score (ACS) differences between five injury subgroups (no injury, brain injury with current symptoms, brain injury without current symptoms, nonbrain injury with current symptoms, and nonbrain injury without current symptoms) using general linear mixed modeling. Almost 11% (70,472/641,285) reported brain injury in the 4 years before assessment. The ACS differed significantly by injury group ($p < 0.0001$). In comparison to the no injury group, those reporting brain injury with current symptoms ($d = -0.44$) and nonbrain injury with current symptoms ($d = -0.24$) demonstrated significantly reduced ACS scores ($p < 0.0001$) indicative of reduced neurocognitive proficiency. In this population-based study of U.S. military personnel, neurocognitive performance was significantly associated with reported injury within the past 4 years among those experiencing current symptoms. Occupational programs focusing on prospective brain health of injured population groups are warranted.

INTRODUCTION

The prospective cognitive and neurological health of military personnel[1] is of considerable concern, in light of the heightened awareness of the health consequences of traumatic brain injury (TBI) events and other experiences occurring in operational and training environments.[2,3] Additionally, the publicity surrounding the high rate of sports-related head injury in high

*Military Performance Division, U.S. Army Research Institute of Environmental Medicine, Kansas Street, Building 42, Natick, MA 01760.
†Research Service, VA Boston Healthcare System, 150 South Huntington Avenue, Boston, MA 02130.
‡Department of Environmental Health, Boston University School of Public Health, 715 Albany Street, Boston, MA 02118.
§Neurocognitive Assessment Branch, U.S. Army Office of The Surgeon General, 3925 Chambers Pass Road, Building 3661, Fort Sam Houston, TX 78234.
‖ Department of Veterans Affairs Center of Excellence for Research on Returning War Veterans, 4800 Memorial Drive, Waco, TX 76711.
¶Cognitive Science Research Center, University of Oklahoma, 3200 Marshall Avenue, Suite 260, Norman, OK 73072.
The views expressed in this article are those of the authors and do not reflect the official policy or position of the Department of Veterans' Affairs, the Department of Defense, or the U.S. Government.
doi: 10.7205/MILMED-D-14-00298

school, collegiate, and professional athletes has served to illuminate and drive research efforts to better understand the long-term effects of brain injury on health and performance.[4] Computer-based cognitive testing programs have been employed as a tool to screen for injury-related changes in cognitive status.[5,6]

In 2008, a Congressionally mandated program was established requiring all Department of Defense (DoD) service members deploying to Iraq or Afghanistan to complete a computer-based neurocognitive assessment.[7] To comply with DoD's clinical testing policy,[8] the Neurocognitive Functional Assessment Program was initiated, which established baseline neurocognitive status of all U.S. service members within 12 months before deployment using the Automated Neuropsychological Assessment Metrics (ANAM; version 4) TBI Military (ANAM4 TBI-MIL) battery. The ANAM4 TBI-MIL is a computer-based set of tests designed to measure cognitive performance across several functional domains, including executive functioning, attention, memory, response time, and information processing speed (Center for the Study of Human Operator Performance ANAM4. TBI-MIL: User Manual. Norman, Oklahoma: Center for the Study of Human Operator Performance; University of Oklahoma,

2007). Previous studies have documented the psychometric properties of ANAM tests[9–11] and normative data for military personnel[12] have been provided. ANAM tests have been shown to be sensitive to the effects of mild brain injury,[13–16] especially in the acute phases[17–19] following injury. More recently, the ANAM test battery has been demonstrated to validly detect impairments in a mixed clinical patient sample.[20]

For this project, we integrated the ANAM4 TBI-MIL data into an analytical database (ANAM4TBI Military Performance Database, AMP-D) to examine neurocognitive performance metrics and factors that may influence performance. Given the emerging focus of brain health as a public health issue worldwide in both military and civilian populations (e.g., Army Performance Triad, The Brain Research through Advancing Innovative Neurotechnologies Initiative, European Year of the Brain[21–23]), knowledge and understanding of the role that particular factors, especially modifiable ones, play in neurocognitive performance is a critical requirement from which appropriate prevention, training, intervention, and treatment programs can be launched.

In this report, we used the AMP-D to examine neurocognitive performance and mood state profiles of DoD personnel completing the ANAM4 TBI-MIL. We compared performance and mood among military personnel who reported having brain or nonbrain injuries in the 4 years before their first ANAM4 TBI-MIL assessment and those reporting no injury. We predicted that having experienced an injury within the past 4 years, particularly where symptoms persist, is associated with reduced neurocognitive proficiency and adverse mood states.

METHODS

The study protocol was reviewed and approved by the Institutional Review Board at the U.S. Army Research Institute of Environmental Medicine and complied with all institutional guidelines for the protection of human subjects.

Study Population

The study population included all U.S. military personnel (n = 671,435) who were administered the ANAM4 TBI-MIL battery starting in 2007 through December 2010 as part of the mandated clinical testing policy.[7,8,24]

Procedures

The ANAM4 TBI-MIL is a battery of tests administered via laptop computer, which takes approximately 20 minutes to complete (Table I). ANAM4 TBI-MIL incorporates two questionnaires requesting demographic and injury information (Demographics, TBI Questionnaire), two questionnaires requiring self-assessment of current state of arousal and mood (Sleepiness Scale [SLP], Mood Scale [MOO]), and seven performance tests (Simple Reaction Time [SRT], Code Substitution-Learning [CDS], Procedural Reaction Time [PRO], Mathematical Processing [MTH], Matching to Sample [M2S], Code Substitution-Delayed [CDD], and Simple Reaction Time Repeated [SR2]). More detailed descriptions of these tests have been provided elsewhere.[9,12,25]

Under the DoD-mandated clinical testing program, ANAM4 TBI-MIL administration was conducted in a standardized manner by trained test proctors at designated sites. The battery was administered primarily in groups, during the daytime hours, in a quiet room. All except two test modules (CDD, SR2) began with practice items to assist in learning the procedures and instructions before the actual test data collection occurred. If a participant did not understand the instructions, test proctors were present to provide clarification and answer questions. Per field operational procedures,

TABLE I. ANAM4 TBI-MIL Battery and Functional Domains Assessed

Task Task List	Abbreviation	Task Description/Functional Domains	Task Parameters Examined
TBI Questionnaire	TBQ	Report of TBI/Other Injury in Past 4 years and Past and Current Symptomatology	Injury Event Type and Related Health Symptoms
Sleepiness Scale	SLP	Assessment of Current Level of Sleepiness	Response Options are Ratings 1–7
Mood Scale	MOO	Assessment of Current Mood State in 7 Categories (7 Subscales: Vigor, Happiness, Depression, Anger, Fatigue, Anxiety, and Restlessness)	Mean of the Adjective Scores for Each Subscale
Simple Reaction Time	SRT	Basic Neural Processing (Motor Activity Speed/Efficiency)	Mean RT, % Correct, Throughput
Code Substitution Learning	CDS	Associative Learning (Speed/Efficiency)	Mean RT, % Correct, Throughput
Procedural Reaction Time	PRO	Processing Speed (Choice Reaction Time Rule Adherence)	Mean RT, % Correct, Throughput
Mathematical Processing	MTH	Working Memory	Mean RT, % Correct, Throughput
Matching to Sample	M2S	Visual Spatial Memory	Mean RT, % Correct, Throughput
Code Substitution-Delayed	CDD	Memory (Delayed)	Mean RT, % Correct, Throughput
Simple Reaction Time (R)	SR2	Basic Neural Processing (Motor Activity Speed/Efficiency)	Mean RT, % Correct, Throughput

(R), Task was repeated at the end of the battery administration to provide a measure of response variation, an indicator of fatigue over the administration time period. RT, Response Time.

data for each test were screened upon completion for potentially invalid test performance (defined as accuracy scores less than or equal to 56%), which could indicate potential misunderstanding of directions or poor effort. Individuals with test performances falling below these accuracy criteria were provided with clarification of the test instructions and asked to repeat that given test.

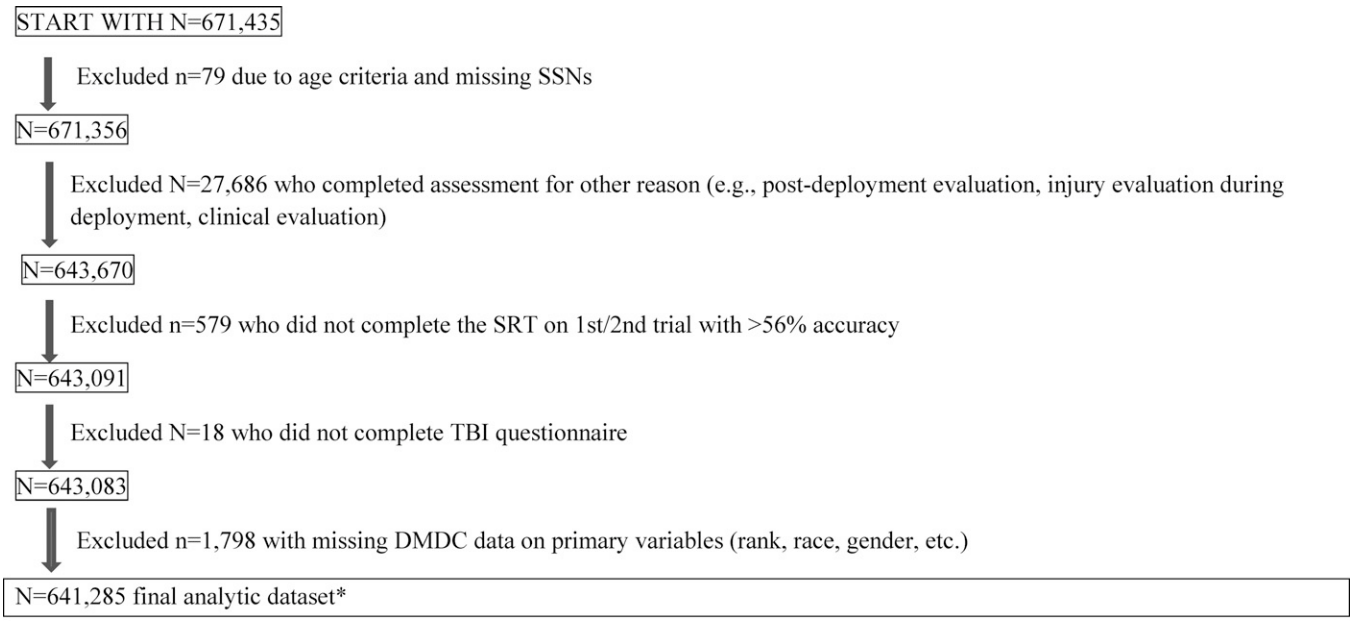
Individual data files of ANAM4 TBI-MIL assessments were obtained from the ANAM Program Office (Neurocognitive Assessment Branch), U.S. Army Office of the Surgeon General. Military service and deployment history data, as well as other demographic (e.g., age, education level, sex, race) and military service (rank, service branch, component, and occupation) information were requested and provided by the Defense Manpower Data Center (DMDC) for use in this project through approved research processes. These data sources were integrated to form the master database (AMP-D) housed and managed at U.S. Army Research Institute of Environmental Medicine.

Data Analyses

In this report, we examined the data from 641,285 individuals administered the ANAM4 TBI-MIL as part of the standard predeployment procedures. This subset includes those individuals 18 to 65 years of age who completed the TBI questionnaire module and at least the SRT test (the first test in the battery) with higher than 56% recorded task accuracy on the first administration (or second in the case of retest) administration and for whom pertinent DMDC personnel data were available. (Exclusions included: 79 due to missing linkage identifiers; 27,686 because they completed the battery for some other reason [such as for a clinical evaluation or postinjury assessment]; 18 who did not complete the TBI questionnaire module; 579 who did not meet SRT task accuracy criteria; and 1,798 because they were missing pertinent DMDC demographic information.) For those individuals who completed ANAM4 TBI-MIL more than once during this period due to multiple deployments between 2007 and 2010 (n = 73,702), only data from the first assessment date were included (Fig. 1).

In addition to the SRT test, all other ANAM4 TBI-MIL performance tests (CDS, M2S, PRO, MTH, CDD, and SR2) were evaluated to determine whether each was completed with greater than 56% accuracy on the first or second test administration within the same calendar day, therefore satisfying test-specific field retest criteria. If a person did not meet the test-specific retest criteria or if test data were missing, their data for that test were not included in the analyses. The percentage of persons excluded by task was as follows: CDS, 0.04%; PRO, 0.15%; MTH, 0.28%; M2S, 0.49%; CDD, 1.35%; and SR2, 0.06%.

The mean, median, and range values for all test-specific scores were computed. Mean response time (mean RT) for correct responses, percentage correct (% correct), and throughput (TP) (correct responses per minute of available response time) were the test parameters selected for analyses of the performance tasks. TP represents a combination of reaction time and accuracy.[26] The Sleepiness Scale responses represent



[*n=73,702 have repeat data; we used their 1st assessment data]

FIGURE 1. Flowchart diagram.

a current rating of sleepiness with possible scores ranging from 1 to 7 (higher number indicates greater sleepiness). For each of the seven Mood subscales, six adjectives are presented along with a response set ranging from "not at all" to "very much" (on a 0 to 6 point scale). The mean of the adjective responses for each of the seven Mood subscales was selected for analysis. Higher values indicate greater endorsement of the mood state dimension.

To provide a measure of overall performance on the ANAM4 TBI-MIL cognitive tests, the ANAM composite score (ACS) was computed by converting TP scores for all tasks in the battery to T-scores relative to an age- and gender-matched normative group.[12,25,27] The ACS is reported in standard deviation units with more negative values indicating poorer overall performance. In addition, the ANAM4 Performance Validity Index (PVI) was computed for each individual. The PVI provides an assessment of valid responding and is computed utilizing the accuracy and RT discrepancy scores from four ANAM4 TBI-MIL tasks: M2S, SRT, PRO, and CDS.[28] The PVI total score ranges from 0 to 48 with higher scores indicating greater likelihood of atypical performance effort. In this report, the recommended cut point score of 10, representing a minimum of 90% specificity in an outpatient sample,[28] was selected as an indicator of questionable performance effort.

Pearson and point biserial correlation coefficients were computed to examine the relationship between TP and age, sex, and education level.

To evaluate whether reporting an injury was associated with reduced cognitive proficiency or adverse mood, individuals were categorized into five injury subgroups (no injury, brain injury with current symptoms, brain injury with no current symptoms, nonbrain injury with current symptoms, and nonbrain injury with no current symptoms) based on their responses on the ANAM4 TBI-MIL questionnaire. Persons were asked "During the past 4 years, have you had any injury (head or other) from any of the following (events)?." Those individuals who did not endorse any injury event in the 4 years before the ANAM4 TBI-MIL assessment comprised the "no injury" group. Individuals were categorized in the "brain injury" group if they reported an injury event in the prior 4 years accompanied by an alteration of consciousness (defined by endorsing at least one of the following symptoms: feeling dazed and confused, experiencing a loss of consciousness, or experiencing loss of memory for the injury or post-traumatic amnesia for the event). Those brain injuries self-reported in the ANAM4 TBI-MIL questionnaire responses were presumed to be mild (rather than meeting moderate or severe classification criteria), as all individuals were actively serving in the military and scheduled for upcoming deployment duty. Detailed data regarding the exact date, type, and severity of injuries were not collected within the ANAM4 TBI-MIL. Those persons who reported an injury in the prior 4 years but did not report alteration or loss of consciousness or loss of memory for the injury event were categorized into the "nonbrain injury" groups. Persons in the two injury subgroups were further classified as reporting injury-related symptoms at the time of testing either at rest or upon exertion (current symptoms) or symptoms only at the time of injury (without current symptoms). By questionnaire design, only those persons endorsing an injury event were then subsequently queried about specific symptoms.

To examine differences in the ACS and mood measures by injury subgroup, linear mixed model analyses were conducted. To evaluate individual injury subgroup differences, adjustment for multiple comparisons with the method of Games–Howell[29] was applied. Additional mixed models were run to examine the ACS and mood measures by injury subgroups while adjusting for sex, age, and education.

Percentile cut scores indicative of below and above aver- age performance (at the 9th and 91st percentile, respec- tively[30]) were calculated for the ACS for the "no injury" group (<1.3 SD below the group mean). Within the four injured groups, the proportion of individuals with below average performance was determined.

A set of post hoc sensitivity analyses was conducted to examine whether questionable performance levels (as determined by the PVI), more severe reported brain injury, or prior deployment influenced ACS differences observed across injury subgroups. Separate linear mixed effect models were conducted, after excluding those persons who (i) met criteria for questionable performance effort or (ii) reported loss of consciousness >20 minutes. We also examined the differences among the injury subgroups stratified by previous deployment history.

All statistical analyses were conducted using SAS (ver- sion 9.3). Because of the large population size, statistical analyses were conducted with significance level $\alpha < 0.001$. Cohen's *d* effect sizes also were computed. For data reduction purposes and to lessen the possibility of Type I error, statistical analyses only examined the ACS rather than each ANAM4 TBI-MIL performance test separately.

RESULTS

The U.S. military population completing ANAM4 TBI-MIL assessments as part of the DoD-wide mandated predeployment program from its onset through the end of December 2010, was on average 28.5 years of age (SD = 7.9) (Table II) at the time of assessment. A total of 64,568 persons (10.1%) were of Hispanic ethnicity. Army personnel made up the largest service branch represented (67%). Almost half (46.2%) of the personnel had deployed previously before the initiation of the DoD-wide Neurocognitive Functional Assessment Program, with the majority of the previous deployments (98%) being to Iraq or Afghanistan since 2001 as part of Operation Iraqi Freedom or Operation Enduring Freedom.

The total number of U.S. military deployed by Fiscal years 2008, 2009, and 2010 was 628, 329, 647, 969, and 623,028,

TABLE II. Characteristics of Those Completing ANAM4 TBI-MIL Battery (n = 641,285)

Variable[a]		*N*	%
Age	<21	92,417	14.41
	21–30 Years	355,511	55.44
	31–40 Years	136,648	21.31
	>40 Years	56,709	8.84
Sex	Male	573,564	89.44
	Female	67,721	10.56
Education	<High School	9,968	1.55
	High School Graduate	531,543	82.89
	College Graduate (4-Year Degree)	69,867	10.89
	Advanced Degree	24,176	3.77
	Unknown	5,731	0.89
Race	White	486,507	75.86
	Black	93,503	14.58
	Asian	19,567	3.05
	American Indian	6,348	1.0
	Other/Unknown	35,360	5.14
Rank	E1–E4	335,661	52.34
	E5–E6	174,301	27.18
	E7–E9	46,779	7.30
	Officer (Includes Warrant)	84,544	13.18
Length of Time in Service, at Time of Assessment	<1 Year	38,364	5.98
	1–<5 Years	276,362	43.10
	5–<10 Years	137,628	21.46
	10+ Years	179,508	27.99
	Unknown (No Service Entry Date)	9,423	1.47
Branch of Service	Army	431,464	67.28
	Air Force	91,802	14.32
	Marine Corps	89,333	13.93
	Navy	28,044	4.37
	Other	642	0.10
Component	National Guard	119,071	18.57
	Regular (Active Duty)	464,686	72.46
	Reserve	57,528	8.97
AFQT/ASVAB Category[b]	I	26,101	4.07
	II	186,247	29.04
	III	312,191	48.68
	IV or V	15,399	2.39
	Unknown or not Available	101,347	15.80

AFQT, Armed Forces Qualification Test; ASVAB, Armed Services Vocational Battery. [a]Data provided by DMDC. [b]AFQT/ASVAB data is primarily only available for enlisted personnel. Being in a lower category (i.e., Category I) indicates better proficiency.

TABLE III. ANAM4 TBI-MIL Battery Performances by Test for Those Completing ANAM4 TBI-MIL Battery

Test	*N*[a]	Variable	Mean (SD)	Median
Simple Reaction Time (SRT)	641,285	Mean RT	264.0 (117.9)	251.8
		% Correct	100.0 (0.6)	100.0
		TP	234.4 (31.6)	238.3
Code Substitution Learning (CDS)	641,031	Mean RT	1154.4 (280.2)	1100.7
		% Correct	97.6 (3.0)	98.6
		TP	53.2 (11.6)	53.2
Procedural Reaction Time (PRO)	640,350	Mean RT	592.0 (107.1)	573.5
		% Correct	96.8 (4.6)	96.9
		TP	100.5 (14.8)	101.7
Mathematical Processing (MTH)	639,518	Mean RT	2821.0 (818.0)	2674.9
		% Correct	93.5 (7.5)	95.0
		TP	21.5 (6.4)	21.1
Matching to Sample (M2S)	638,175	Mean RT	1693.6 (482.7)	1624.2
		% Correct	94.5 (6.6)	95.0
		TP	35.6 (10.9)	34.5
Code Substitution Delayed (CDD)	632,586	Mean RT	1245.3 (370.5)	1168.8
		% Correct	91.3 (9.5)	94.4
		TP	46.5 (15.8)	46.1
Simple Reaction Time (SR2)	640,912	Mean RT	263.7 (82.2)	251.4
		% Correct	100 (0.4)	100.0
		TP	234.5 (32.7)	238.6

[a]Includes those individuals who completed the test on the first (or second in the case of retest) administration with >56% accuracy. RT, Response time; TP, Throughput.

respectively (data report from DMDC, written communication, January 2013). Compared to the deployed U.S. military population serving in 2009, those completing the ANAM4 TBI-MIL during 2007 to 2010 were similar in terms of sex, race/ethnicity, and service component characteristics. They were somewhat more likely to be from the lower enlisted ranks and in Army service than the U.S. deployed population in 2009 (which was approximately 60% Army, 16% Air Force, 10% Marine Corps, and 15% Navy, with 42% from enlisted E1 to E4 ranks).

The mean, standard deviation, and median values for each test are presented in Table III. Less than 2% (1.72%) of the 634,155 persons for whom the PVI was able to be computed met criteria for questionable performance effort ("no injury" group, 1.45%; "brain injury with current symptoms" group, 5.92%; "brain injury without current symptoms" group, 2.07%; "nonbrain injury with current symptoms" group, 3.84%; and "nonbrain injury without current symptoms" group, 1.86%). The mean PVI score overall was 1.29 (SD = 2.51; standard error of mean = 0.003).

The correlations between TP, age, gender, and education were statistically significant for all tests ($p < 0.0001$) (Table IV). For age, all Pearson correlation coefficients were negative with the exception of MTH, which was positive ($r = 0.139$), indicating that older persons performed better on

TABLE IV. Correlations Between ANAM4 TBI-MIL Test Throughput and Demographic Characteristics

Test	Age[a]	Gender[b]	Education Level[b]
SRT	− 0.169	− 0.085	− 0.021
CDS	− 0.291	− 0.028	− 0.051
PRO	− 0.148	− 0.019	0.021
MTH	0.139	0.005	0.229
M2S	− 0.178	− 0.099	0.004
CDD	− 0.284	− 0.034	− 0.032
SR2	− 0.126	− 0.087	0.004

[a]Pearson's correlation coefficients. [b]Point biserial correlation coefficients (Sex [M = 0/F = 1]; Education [HS or less = 0/>HS = 1]). All correlation coefficients significant at $p < 0.0001$.

TABLE V. Description of Injury Groups Completing the ANAM4 TBI-MIL Battery Predeployment

			Brain Injury in Prior 4 Years		Nonbrain Injury in Prior 4 Years	
	ALL	No Injury in Prior 4 Years	With Current Symptoms	Without Current Symptoms	With Current Symptoms	Without Current Symptoms
N (% of total)	641,285	521,605 (81.3)	25,349 (4.0)	45,123 (7.0)	7,952 (1.2)	41,256 (6.4)
Male, *N* (%)	573,564 (89.4)	463,357 (88.8)	23,886 (94.2)	42,127 (93.3)	7,080 (89.0)	37,114 (89.9)
Age (Mean [SD])	28.5 (7.9)	28.7 (8.0)	27.8 (6.7)	27.1 (6.9)	28.9 (7.5)	28.1 (7.5)
>High School, *N* (%)	94,043 (14.7)	81,819 (12.8)	1,576 (0.2)	4,600 (0.7)	701 (0.1)	5,347 (0.8)
Previously Deployed, *N* (%)	296,472 (46.2)	229,236 (44.0)	17,487 (69.0)	23,809 (52.8)	5,138 (64.6)	20,802 (50.4)
Injury Scenario[a]:						
Blast, *N* (%)			12,849 (50.7)	8,353 (18.5)	2,639 (33.2)	4,157 (10.1)
Bullets, *N* (%)			137 (0.5)	111 (0.3)	24 (0.3)	98 (0.2)
Fragment, *N* (%)			845 (3.3)	464 (1.0)	108 (1.4)	237 (0.6)
Vehicular, *N* (%)			8,405 (33.2)	11,801 (26.2)	1,961 (24.7)	11,592 (28.1)
Sports, *N* (%)			5,250 (20.7)	11,766 (26.1)	2,054 (25.8)	12,720 (30.8)
Fall, *N* (%)			7,440 (29.4)	10,767 (23.9)	2,049 (25.8)	8,502 (20.6)
Fight, *N* (%)			4,894 (19.3)	9,338 (20.7)	903 (11.4)	6,109 (14.8)
Other Blow, *N* (%)			6,414 (25.5)	9,998 (22.3)	999 (12.6)	4,563 (11.1)
ANAM4 Sleep Scale (Mean [SD])	2.22 (1.11)	2.17 (1.07)	2.94 (1.35)	2.40 (1.19)	2.66 (1.33)	2.24 (1.13)
ANAM4 Composite Score[b] (Mean[SEM])	0.073 (0.001)	0.096 (0.001)[c]	− 0.347 (0.008)	0.054 (0.005)	− 0.167 (0.013)	0.095 (0.006)

SD, Standard deviation; SEM, Standard error of mean. [a]Responders can select more than one category so values will not add to 100%. [b]Higher values indicate better performance. Sample size for ANAM4 Composite Score, n = 627,887 ("no injury," n = 511,038; "head injury with current symptoms," n = 24,588; "brain Injury without current symptoms," n = 44,197; "nonbrain injury with current symptoms," n = 7,735; "nonbrain injury without current symptoms," n = 40,329). [c]Effect sizes for difference between "no injury" and the four groups were − 0.44 ("brain injury with current symptoms"), − 0.04 ("brain Injury without current symptoms"), − 0.26 ("nonbrain injury with current symptoms"), and 0.001 ("nonbrain injury without current symptoms").

MTH. For sex, the point biserial correlations were all negative and <− 0.10, except for MTH (r = 0.005). The point biserial correlations between education level and TP all tended to fall around zero and demonstrated a mixed pattern, where having a college or advanced degree was positively correlated with PRO, MTH, M2S, and SR2 but negatively correlated with SRT, CDS, and CDD.

Almost 11% of the population reported having a brain injury in the 4 years before the assessment (Table V) and 7% reported incurring exclusively a nonbrain injury in the previous 4 years. The most prevalent mechanism resulting in the injury reported by either the "brain injury with current symptoms" or "nonbrain injury with current symptoms" groups was blast (50.7% and 33.2% respectively). Among the groups reporting "brain injury without current symptoms", the most prevalent injury mechanisms reported were vehicular (26.2%) or sports (26.1%), while injury during sports (30.8%) was most prevalent among the "nonbrain injury without current symptoms" group.

Approximately 50% of the "brain injury with current symptoms" group reported some loss of consciousness at the time of their injury, with headaches (67.2%) and ringing in the ears (47.2%) being the most prevalent symptoms reported as being present at the time of their injury (Table VI). Similarly, for the "nonbrain injured with current symptoms" group, headaches (28.2%) and ringing in the ears (20.2%) were the most prevalent symptoms reported at the time of injury. With respect to current symptoms, the two most prevalent symptoms in both groups were sleep problems (51.4% in brain injured and 34.8% in nonbrain injured) and irritability or short temper (49.9% in brain injured and 31.2% in nonbrain injured).

The ACS differed significantly by injury group (F [4, 627886] = 1180.58, p < 0.0001) (Table V). No significant difference in ACS between the "no injury" and the "nonbrain injury without current symptoms" groups was observed. The "brain injury with current symptoms" group demonstrated a significantly reduced ACS indicating reduced proficiency compared to the "nonbrain injury with current symptoms" group. In turn, both injury groups with current symptoms recorded significantly lower ACSs compared to the "brain injury without current symptoms" group.

Figure 2 presents the cumulative frequency distributions of the ACS for the "no injury" and "brain injury with cur- rent symptoms" groups. The medians (50th percentiles) of the two groups differ by 0.3 ("no injury:" 0.12 [SD 1.0; variance 1.03]); "brain injury with current symptoms:" − 0.18 (SD 1.36; variance 1.84). At the lower tail of the distribution for the ACS, the "brain injury with current symptoms" group (21%) was two times more likely and the nonbrain injury with current symptoms" group (16%) was one and a half times more likely than the "no injury" group (9%) to perform in the below average range for the ACS.

The ANAM4 TBI-MIL Sleep score (Table V) significantly differed by injury group (F [4,641,159] = 3708.16, p < 0.0001), with all injury groups showing significant differences from each other.

For each of the mood state subscales, significant differences (all p < 0.0001) were observed by injury group (n = 641,275). The pattern of results was similar for each subscale, in that

TABLE VI. Description of Symptoms Reported by Injury Groups Completing the ANAM4TBI-MIL Task Battery at Predeployment

	Reporting Brain Injury in the Prior 4 Years		Reporting Nonbrain Injury in the Prior 4 Years	
	With Current Symptoms	Without Current Symptoms	With Current Symptoms	Without Current Symptoms
Symptoms	*N* = 25,349	*N* = 45,123	*N* = 7,952	*N* = 41,256
Dazed, Confused, Saw Stars (%)	86.6	78.7		
Knocked out— <1 Minute (%)	31.2	27.6		
Knocked out— From 1–20 Minutes (%)	14.8	10.0		
Knocked out— >20 Minutes (%)	3.1	2.1		
Did not Remember Injury (%)	16.6	11.7		
Headaches (%)				
At Time of Injury	67.2	61.4	28.2	21.7
Currently[a]	48.5	—	30.7	—
Nausea/Vomiting (%)				
At Time of Injury	20.7	12.5	3.8	2.7
Currently	6.8	—	3.8	—
Sensitivity to Bright Light/Noise (%)				
At Time of Injury	33.9	17.8	8.4	3.4
Currently	28.6	—	14.1	—
Balance Problems/Dizziness (%) At				
Time of Injury	38.5	28.9	8.8	5.0
Currently	23.4	—	11.8	—
Ringing in the Ears (%) At				
Time of Injury	47.2	26.7	20.2	7.2
Currently	41.2	—	26.5	—
Sleep Problems (%) At				
Time of Injury	35.1	11.9	15.8	4.2
Currently	51.4	—	34.8	—
Irritability (Short Temper) (%) At				
Time of Injury	30.2	10.7	11.6	3.3
Currently	49.9	—	31.2	—
Memory Problems/Lapses (%) At				
Time of Injury	33.2	13.5	8.2	2.3
Currently	49.4	—	26.1	—
Other Symptoms (%) At				
Time of Injury	8.7	4.5	19.6	7.1
Currently	10.5	—	29.1	—

[a]Defined as reporting symptom currently either while resting or upon exertion.

all injury groups differed from each other (Fig. 3). Both the "brain injury with current symptoms" followed by the "nonbrain injury with current symptoms" groups consistently endorsed significantly higher symptoms of restlessness, fatigue, anger, depression, and anxiety than the other three groups, whereas the "no injury" group reported significantly more positive feelings of vigor and happiness compared to the injury groups.

After accounting for age, sex, or education differences among the injury subgroups, there was no difference in the pattern of significant results observed for the ACS, Sleep scale, and Mood subscales. In posthoc analyses, the pattern of significant results for the ACS by injury subgroup did not differ following exclusion of those persons who met criteria for questionable effort based on the PVI or exclusion of the subset of the brain-injured groups that reported loss of consciousness greater than 20 minutes. Also, the pattern of results was similar when stratified by previous deployment history: among the deployed subgroups, moderate effect sizes were observed when comparing the differences between the "no injury" group and the "brain injury with current symptoms" ($d = -0.41$) group and the "nonbrain injury with current symptoms" ($d = -0.51$) group.

DISCUSSION

The population-based AMP-D represents the first available research resource to enable the examination of neurocognitive profiles of the U.S. military population. Our cross-sectional results demonstrate the association between neurocognitive performance and reported prior injury, particularly among those who continue to experience symptoms, with the group reporting "brain injury with current symptoms" followed by the group reporting "nonbrain injury with current symptoms" demonstrating reduced proficiency (as assessed by the ANAM4 TBI-MIL Composite score) compared to those groups reporting "brain or nonbrain injury with no current symptoms or no injury." Compared to those reporting "no injury," military personnel reporting "brain injury with

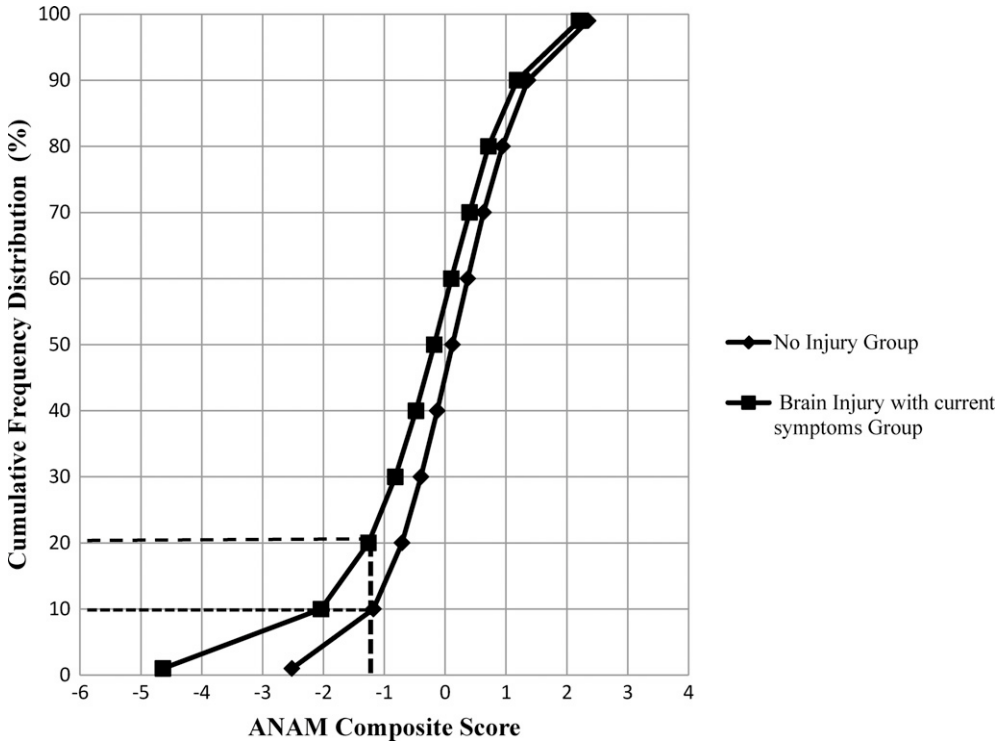


FIGURE 2. Cumulative frequency distribution of the ANAM Composite Score in persons in the "no injury" and "brain injury with current symptoms" groups.

current symptoms" were two times more likely to function at below average levels.

It is important to comment that what is statistically significant may not always translate into meaningful differences in biological or clinical terms.[31] However, the observed moderate effect size magnitude for the difference in neurocognitive proficiency on the ANAM4 TBI-MIL battery between the "no injury" group and the "brain injury with current symptoms" group (d = 0.44), does suggest a clinically meaningful result; an effect to a lesser degree was found with the "nonbrain injured with current symptoms" (d = 0.24) group. From a population-based public health perspective, even a small magnitude shift in the group distribution toward poorer neurocognitive proficiency (indicative of subtle population shifts) may be widely relevant. Although less severe than observed in landmark studies by Needleman and colleagues in the late 1970s and early 1980s documenting a threefold difference in IQ levels <80 among high-lead exposed children,[32,33] our results represent a similar implication at the population level. Even a small shift in the performance mean of the popu- lation related to prior (brain) injury with current symp- toms could be viewed as a benchmark of change, which over time may lead to increasing number of individuals

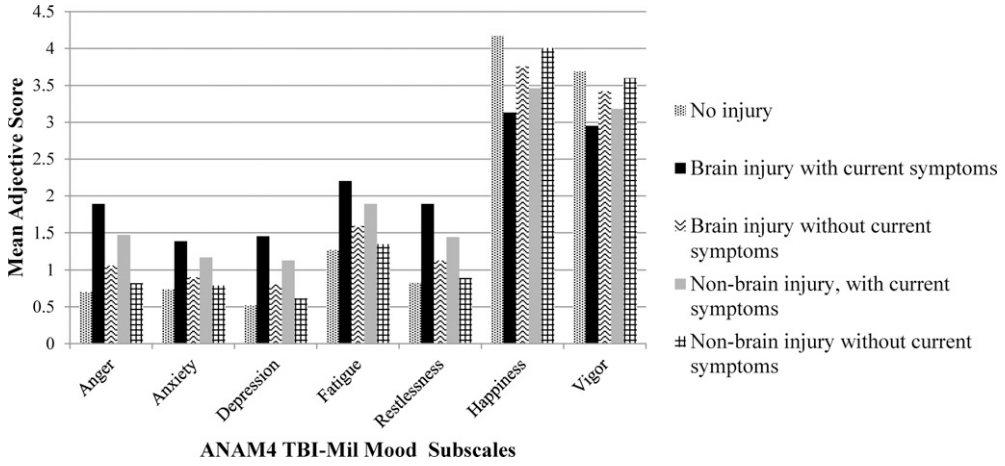


FIGURE 3. Mood states reported by injury groups completing the ANAM4 TBI-MIL task battery at predeployment.

seeking care and connote significant implications for public health policy.

Previously reported results have been inconsistent as to whether history of concussion is associated with poorer cognitive proficiency on computer-assisted cognitive testing[34–37] but prior studies did not further compare head injured groups with and without current symptoms. In this study, we observe significantly reduced cognitive proficiencies among those persons with a reported history of previous injury (brain and nonbrain) and who are currently experiencing symptoms. The temporal nature and number of injury events in relation to the time of assessment, injury classification criteria followed, and group-level factors (such as degree of effort or motivation, and role of other health comorbidities) may contribute to the differences in the results observed in this study. Our study addressed injury within a temporal window of the previous 4 years (rather than at any time in the past) and utilized brain/head injury criteria defined by consensus-based symptom reports and validated against a clinical interview diagnostic approach.[38] Less than 2% of the population met criteria for questionable effort, which is decidedly lower than 6.3–27.9% reported by other studies[39,40] involving computer-based neurocognitive assessment of brain injury between high school and college athletes, albeit using different metrics.

There are a number of strengths to this study. By the structure of the AMP-D, this study permits the descriptive analyses of predeployment cognitive assessment of the total population of the U.S. military. Therefore, findings represent those of all deployed military, independent of health care–seeking behavior or other sampling biases. The reliance on self-report of injury events in a retrospective manner, in particular those specifics related to the severity and temporal nature of the injury related to the time of assessment, presents a limit to the study. Review of associated medical records pertaining to specific injury events may provide additional details, but it is important to note that not all injuries may result in the individual seeking clinical care. The current analytic framework of the AMP-D does not include the ability to address the role(s) of multiple potential confounding factors, such as comorbid mental and physical health conditions. Future analytic steps will integrate clinical medical encounter diagnostic data (for a population subset) into the AMP-D and thus enable analysis of the role of comorbid disorders, such as post-traumatic stress disorder and major depression, on performance. In addition, concordance analyses between reported brain and nonbrain injuries in ANAM4 TBI-MIL assessment and those injury events documented within the clinical healthcare system are planned.

The influence of injury, not just brain injury, on neurocognitive performance over one's military career and in the years following service, warrants continued attention. The AMP-D resource fills a critical capability gap permitting the evaluation of population-based brain health and performance trends and examination of both positive and protective factors and adverse risk elements that may influence performance.

ACKNOWLEDGMENTS

We thank the staff at the U.S. Army Office of the Surgeon General, Neurocognitive Assessment Branch, as well as the DoD Defense Manpower Data Center for their support in this project. Funding for this project has been provided by the U.S. Army Medical Research and Materiel Command to the U.S. Army Research Institute of Environmental Medicine and through award #W81XWH-08-1-0021 (PI: SP Proctor) to the Henry M. Jackson Foundation for the Advancement of Military Medicine, Inc.

REFERENCES

1. Vasterling JJ, Proctor SP: Understanding the neuropsychological consequences of deployment stress: a public health framework. J Int Neuropsychol Soc 2011; 17(1): 1–6.
2. Okie S: Traumatic brain injury in the war zone. N Engl J Med 2005; 352(20): 2043–47.
3. Committee on Gulf War and Health. Brain Injuries in Veterans and Long-term Health Outcomes: Gulf War and Health: Volume 7: Long-term Consequences of Traumatic Brain Injury. Washington, DC, The National Academies Press, 2009.
4. Corrigan JD, Hammond FM: Traumatic brain injury as a chronic health condition. Arch Phys Med Rehabil 2013; 94(6): 1199–201.
5. Guskiewicz KM, Bruce SL, Cantu RC, et al: Recommendations on management of sports-related concussion: summary of the National Athletic Trainers' Association position statement. Neurosurg 2004; 55(4): 891–5.
6. Guskiewicz KM, Bruce SL, Cantu RC, et al: National Athletic Trainers' Association position statement: management of sport-related concussion. J Athl Train 2004; 39(3): 280–97.
7. US Government Accountability Office, Report to Congressional Addresses: DoD Health Care Mental Health and Traumatic Brain Injury Screening Efforts Implemented, but Consistent Pre-deployment Medical Record Review Policies Needed. GAO-08-615. Washington, DC, US GAO, 2008. Available at http://www.gao.gov/products/gao-08-615; accessed February 28, 2014.
8. Department of Defense, Health Affairs: Memorandum on Baseline Predeployment Neurocognitive Functional Assessment. Washington, DC, Health Affairs, 2008. Available at http://www.health.mil/~/media/MHS/Policy%20Files/Import/baseline_pre-deployment_neurocognitive_functional_assessment_-_interim_guidance.ashx; accessed February 28, 2014.
9. Vincent AS, Bleiberg J, Yan S, et al: Reference data from the Automated Neuropsychological Assessment Metrics for use in traumatic brain injury in an active duty military sample. Mil Med 2008; 173: 836–52.
10. Roebuck-Spencer TM, Reeves DL, Bleiberg J, et al: Influence of demographics on computerized cognitive testing in a military sample. Mil Psychol 2008; 20: 187–203.
11. Kaminski TW, Groff RM, Glutting JJ: Examining the stability of Automated Neuropsychological Assessment Metric (ANAM) baseline test scores. J Clin Exp Neuropsychol 2009; 31: 689–97.
12. Vincent AS, Roebuck-Spencer T, Gilliland K, Schlegel R: Automated Neuropsychological Assessment Metrics (v4) Traumatic Injury Battery: military normative data. Mil Med 2012; 177(3): 256–69.
13. Warden DL, Bleiberg J, Cameron KL, et al: Persistent prolongation of simple reaction time in sports concussion. Neurology 2001; 57(3): 524–6.
14. Bleiberg J, Cernich A, Cameron K, et al: Duration of cognitive impairment after sports concussion. Neurosurg 2004; 54(5): 1073–8.
15. Sim A, Terrberry-Spohr L, Wilson KR: Prolonged recovery of memory functioning after mild traumatic brain injury in adolescent athletes. J Neurosurg 2008; 108(3): 511–6.

16. Norris JN, Carr W, Herzig T, Labrie W, Sams R: ANAM4 TBI reaction time-based tests have prognostic utility for acute concussion. Mil Med 2013; 178(7): 767–74.
17. Bryan C, Hernandez AM: Magnitudes of decline on Automated Neuropsychological Assessment Metrics subtest scores relative to predeployment baseline performance among service members evaluated for traumatic brain injury in Iraq. J Head Trauma Rehabil 2012; 27(1): 45–54.
18. Coldren RL, Russell ML, Parish RV, Dretsch M, Kelly MP: The ANAM lacks utility as a diagnostic or screening tool for concussion more than 10 days following injury. Mil Med 2012; 177(2): 179–83.
19. Kelly MP, Coldren RL, Parish RV, Dretsch MN, Russell ML: Assessment of acute concussion in the combat environment. Arch Clin Neuropsychol 2012; 27(4): 375–88.
20. Woodhouse J, Heyanka DJ, Scott J, et al: Efficacy of the ANAM General Neuropsychological Screening battery (ANAM GNS) for detecting neurocognitive impairment in a mixed clinical sample. Clin Neuropsychol 2013; 27(3): 376–85.
21. Army Times: Pathway to a Fit and Healthy Force Improving Performance, Resilience, and Readiness in the Army. Available at http://armymedicine.mil/Pages/performance-triad.aspx; accessed March 14, 2014.
22. Baker M: European Year of the Brain 2014: a new impulse to strengthen the alliance for brain health. Croat Med 2013; 54(5): 417–8.
23. Markoff J: Obama seeking to boost study of human brain. New York Times. February 17, 2013. Available at http://www.nytimes.com/2013/02/18/science/project-seeks-to-build-map-of-human-brain.html; accessed January 29, 2014.
24. Department of Defense: Comprehensive Policy on Neurocognitive Assessments by Military Services. Number 6490.13. Department of Defense, 2013. Available at http://www.dtic.mil/whs/directives/corres/pdf/649013p.pdf; accessed February 28, 2014.
25. Vincent AS, Roebuck-Spencer T, Lopez MS, et al: Effects of military deployment on cognitive functioning. Mil Med 2012; 177(3): 248–55.
26. Thorne DR: Throughput: a simple performance index with desirable characteristics. Behav Res Methods 2006; 38: 569–73.
27. Roebuck-Spencer T, Vincent AS, Twillie DA, et al: Cognitive change associated with self-reported mild traumatic brain injury sustained during the OEF/OIF conflicts. Clin Neuropsychol 2012; 26(3): 474–89.
28. Roebuck-Spencer T, Vincent AS, Gilliland K, Johnson DR, Cooper DB: Initial clinical validation of an embedded effort measure within the Automated Neuropsychology Assessment Metrics (ANAM). Arch Clin Neuropsych 2013; 28(7): 700–10.
29. Games PA, Howell JF: Pairwise multiple comparison procedures with unequal n's and/or variances: a Monte Carol study. J Educ Behav Stat 1976; 1(2): 113–25.
30. Hannay HJ, Lezak MD: The neuropsychology examination: interpretation. In: Neuropsychology Assessment, pp 113–156. Edited by Lezak MD, Howieson DB, Loring DW. New York, Oxford University Press, 2004.
31. Todd KH: Clinical versus statistical significance in the assessment of pain relief. Ann Emergency Med 1996; 27(4): 439–41.
32. Needleman HL, Leviton A, Bellinger D: Lead-associated intellectual deficits. N Engl J Med 1982; 306(6): 367.
33. Bellinger DC: What is an adverse effect? A possible resolution of clinical and epidemiological perspectives on neurobehavioral toxicity. Environ Res 2004; 95(2004): 394–405.
34. Matser JT, Kessels AG, Lezak MD, Troost J: A dose-response relation of headers and concussions with cognitive impairment in professional soccer players. J Clin Exp Neuropsychol 2001; 23: 770–4.
35. Broglio SP, Ferrara MS, Piland SG, Anderson RB: Concussion history is not a predictor of computerized neurocognitive performance. Br J Sports Med 2006; 40: 802–5.
36. Ivins BJ, Kane R, Schwab KA: Performance on the Automated Neuropsychological Assessment Metrics in a nonclinical sample of soldiers screened for mild TBI after returning from Iraq and Afghanistan: a descriptive analysis. J Head Trauma Rehabil 2009; 24(1): 24–31.
37. Hutchinson M, Comper P, Mainwaring L, Richards D: Normative data in a sample of Canadian university athletes using ANAM tests. J Clin Sport Psychol 2012; 6: 336–50.
38. Schwab KA, Ivins B, Cramer G, et al: Screening for traumatic brain injury in troops returning from deployment in Afghanistan and Iraq: initial investigation of the usefulness of a short screening tool for traumatic brain injury. J Head Trauma Rehabil 2007; 22: 377–89.
39. Schatz P, Moser RS, Solomon GS, Ott SD, Karpf R: Prevalence of invalid computerized baseline neurocognitive test results in high school and collegiate athletes. J Athl Train 2012; 47(3): 289–96.
40. Szabo A, Alosco ML, Fedor A, Gunstad J: Invalid performance and the impact in national collegiate athletic association division I football players. J Athl Train 2013; 48(6): 851–5.